

Citation 2

A2

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-098707

(43)Date of publication of application : 11.04.1995

(51)Int.Cl.

G06F 17/27

(21)Application number : 05-243146

(71)Applicant : FUJI XEROX CO LTD

(22)Date of filing : 29.09.1993

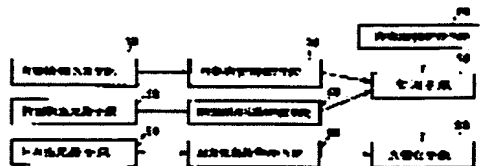
(72)Inventor : KUROSAWA HIROSHI

## (54) DOCUMENT PROCESSOR

## (57)Abstract

**PURPOSE:** To provide a document processor which can output a single structured document or plural document parts forming a single structured document out of a single document whose contents are not structured.

**CONSTITUTION:** A contents information input means 10 supplies the data as the contents information on a document, and a contents information interpreting means 20 interprets the data supplied from the means 10. A logical structure defining means 30 defines the logical structure to be produced, and a logical structure definition interpreting means 40 interprets the contents of definition. An output destination defining means 50 defines the output destination of the document, and an output destination definition interpreting means 60 interprets the contents of definition. A converter means 70 reads the contents information out of a contents information holding means 80 based on the interpreting result of the means 20 and also converts the contents information into the document parts based on the interpreting result of the means 40. A document forming means 90 outputs the converted document parts or a structured document consisting of these document parts to the output destination, i.e., the retrieving result of the means 60.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-98707

(43) 公開日 平成7年(1995)4月11日

(51) Int.Cl. <sup>4</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/27		7315-5L	G 0 6 F 15/ 20	5 5 0 F

審査請求 未請求 請求項の数 6 O L (全 36 頁)

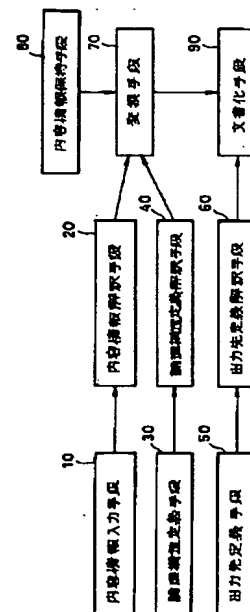
(21) 出願番号	特願平5-243146	(71) 出願人	000005496 富士ゼロックス株式会社 東京都港区赤坂三丁目3番5号
(22) 出願日	平成5年(1993)9月29日	(72) 発明者	黒澤 宏 神奈川県川崎市高津区坂戸3丁目2番1号 K S P R & D ビジネスパークビル 富士ゼロックス株式会社内
		(74) 代理人	弁理士 木村 高久

(54) 【発明の名称】 文書処理装置

(57) 【要約】

【目的】 文書内容が構造化されていない1つの文書から、1つの構造化文書あるいは1つの構造化文書を構成する複数の文書部品を出力することのできる文書処理装置を提供する。

【構成】 内容情報入力手段10は内容情報となるデータを入力し、内容情報解釈手段20はそのデータを解釈する。論理構造定義手段30は生成すべく論理構造の定義し、論理構造定義解釈手段40はその定義内容を解釈する。出力先定義手段50は文書の出力先を定義し、出力先定義解釈手段60はその定義内容を解釈する。変換手段70は、内容情報解釈手段20の解釈結果に基づいて内容情報保持手段80から内容情報を読み込むと共に、この内容情報を、論理構造定義解釈手段40の解釈結果に従って文書部品に変換する。文書化手段90は、変換された文書部品あるいは変換された複数の文書部品を1つの構造化文書として、出力先定義解釈手段60の解釈結果である出力先へ出力する。



## 【特許請求の範囲】

【請求項1】文書内容が構造化されていない文書を保持する保持手段と、

前記文書を複数の部分構造に分割するための情報を示す分割情報を指定する指定手段と、

前記文書を前記分割情報に基づいて複数の部分構造に分割し、当該各部分構造それぞれに、少なくとも、前記文書における論理的なレイアウト上の位置に対応する構造情報を付加する処理手段と、

予め設定された指示に応じて、前記処理手段によって構造情報を付加された複数の部分構造を、それぞれ文書部品として出力するか或いは論理構造を有する1つの文書として出力する文書化手段とを具備したことを特徴とする文書処理装置。

【請求項2】前記指定手段は、前記分割情報に従って分割される複数の部分構造をそれぞれ段落構造とする旨を指定すると共に、当該複数の段落構造に対して適用されるべく段落属性を指定し、

前記処理手段は、前記分割情報に従って分割した複数の段落構造の内容に対して前記段落属性を適用させると共に、当該各段落構造それぞれに前記構造情報を付加することを特徴とする請求項1の文書処理装置。

【請求項3】前記指定手段は、前記分割情報に従って分割される複数の部分構造をそれぞれ段落構造として指定すると共に、当該複数の段落構造に対して適用されるべく予め設定された段落属性を表しているひな型情報を指定し、

前記処理手段は、前記分割情報に従って分割した複数の段落構造の内容に対して前記ひな型情報に対応する段落属性を適用させると共に、当該各段落構造それぞれに前記構造情報を付加することを特徴とする請求項1の文書処理装置。

【請求項4】文書内容が構造化されていない複数の文書を保持する保持手段と、

前記複数の文書毎に複数の部分構造に分割するための情報を示す分割情報を指定すると共に、前記複数の文書間の階層的な関係を定義するための情報を示すレベル情報を指定する指定手段と、

前記文書毎に該当する文書を前記分割情報に基づいて複数の部分構造に分割すると共に、当該複数の文書の分割された複数の部分構造それぞれに、前記レベル情報と各文書の論理的なレイアウト上の位置に対応する構造情報を付加する処理手段と、

予め設定された指示に応じて、前記処理手段によって構造情報を付加された複数の部分構造を、それぞれ文書部品として出力するか或いは論理構造を有する1つの文書として出力する文書化手段とを具備したことを特徴とする文書処理装置。

【請求項5】前記指定手段は、前記分割情報に従って分割される複数の部分構造をそれぞれ段落構造として指定

すると共に、当該複数の段落構造に対して適用されるべく段落属性を指定し、

前記処理手段は、前記分割情報に従って分割した複数の段落構造の内容に対して前記段落属性を適用させると共に、当該各段落構造それぞれに前記レベル情報及び構造情報を付加することを特徴とする請求項4の文書処理装置。

【請求項6】前記指定手段は、前記分割情報に従って分割される複数の部分構造をそれぞれ段落構造として指定すると共に、当該複数の段落構造に対して適用されるべく予め設定された段落属性のひな型情報を指定し、前記処理手段は、前記分割情報に従って分割した複数の段落構造の内容に対して前記ひな型情報に対応する段落属性を適用させると共に、当該各段落構造それぞれに前記レベル情報及び構造情報を付加することを特徴とする請求項4の文書処理装置。

## 【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、テキストファイルから特定の文書フォーマットに適合する文書を生成する文書処理装置に関する。

【0002】

【従来の技術】従来の文書処理装置としては、特開昭62-229364号公報、特開昭63-286963号公報、特開平2-28868号公報に開示されたものが知られている。

【0003】特開昭62-229364号公報のものは、見出し語辞書及び見出し語規則辞書を参照して、入力された文書データから見出し文を抽出し、この抽出結果について文書構造規則を適用して文書データの文書構造を求め、この求められた文書構造の文書データをレイアウト規則に従ってレイアウト処理を施して文書を得るようにしたものである。これにより、所定のレイアウト形式に従った文書を出力可能にしている。

【0004】特開昭63-286963号公報のものは、変換前のテキスト文書から章、節等の見出しを検出し、この見出し構造に制御データ及び管理構造を付加するようにしたものである。これにより、見出しの削除・移動・追加等の操作を容易にしている。

【0005】特開平2-28868号公報のものは、論理構造を持つ文書から抽出された論理構造及び書式情報、又は直接指定された論理構造及び書式情報を書式定義ファイルとして登録する手段を備え、文書の書式設定を行う際、所望の書式を有する文書が、書式定義ファイル中に登録されていれば、その論理構造及び書式情報を書式設定の対象である文書に適用し、一方、登録されていなければ、論理構造及び書式情報を直接設定するようにしたものである。これにより、既に作成された文書から論理構造及び書式情報を抽出し、これらの情報を他の既存の文書又は新規の文書に適用可能にしている。

3

4

【0006】

【発明が解決しようとする課題】上述した各公報に開示されたものは、あくまでも作成された文書に含まれる、章、節、項などの見出しを検出し、この検出結果に基づいて、1つの文書を別のフォーマットの文書に変換するものである。この場合、変換結果は1つの文書である。\*

『

第2章 タイトル

第1節 小見出し

ああああああああ

ああああああああ。

【0008】

iiiiiiiiiiiiiiii

iiiiiiiiiiiiiiii。

【0009】

ここで、  
『』内の内容は文書

『

第2章 タイトル

』

と

『

第1節 小見出し

』

と

『

ああああああああ

ああああああああ。

【0010】

iiiiiiiiiiiiiiii

iiiiiiiiiiiiiiii。

【0011】

』

という3個の部分構造が生成されていた。

【0012】このように従来においては、本文段落となる部分は、上記部分構造2cの様に1つの塊（本文段落構造）として生成されてしまう。★

『

ああああああああ

ああああああああ。

【0014】

』

と

『

iiiiiiiiiiiiiiii

iiiiiiiiiiiiiiii。

【0015】

』

とに分割処理するということである。

\* しかも、上記見出しが存在しない文書を複数の部分構造（段落構造）に分割したり、前記見出しが存在しない複数の文書（複数のテキスト）を1つの文書に構造（段落構造）を付加して変換することはできなかった。

【0007】ここで、上記の問題点を具体例を挙げて説明する。例えば、

…文書1

※「ああ…あ。」および「いい…い。」は文字列（テキスト）という1つの文書1を処理すると、

…部分構造2 a

…部分構造2 b

…部分構造2 c

★【0013】しかし、実際の文書処理においては、上記部分構造2cも複数の段落構造に分割されることが良い場合もある。例えば、上記部分構造2cを更に、

…段落構造3 a

…段落構造3 b

50 【0016】これによって、例えば段落構造3aのみ

を、複写、移動などの編集処理を施すことが可能となり、文書処理がより柔軟になる。

【0017】しかしながら、上記従来においては、本文段落を複数の段落構造に自動的に分割することはできなかったため、上述したような柔軟な文書処理を行うことができない。

【0018】そこで、本発明の目的は、文書内容が構造化されていない1つの文書から、1つの構造化文書あるいは1つの構造化文書を構成する複数の文書部品を出力することのできる文書処理装置を提供することにある。

【0019】また本発明の他の目的は、文書内容が構造化されていない複数の文書から、1つの構造化文書を出力することのできる文書処理装置を提供することにある。

【0020】

【課題を解決するための手段】第1の発明の文書処理装置は、文書内容が構造化されていない文書を保持する保持手段と、前記文書を複数の部分構造に分割するための情報を示す分割情報を指定する指定手段と、前記文書を前記分割情報に基づいて複数の部分構造に分割し、当該各部分構造それぞれに、少なくとも、前記文書における論理的なレイアウト上の位置に対応する構造情報を付加する処理手段と、予め設定された指示に応じて、前記処理手段によって構造情報を付加された複数の部分構造を、それぞれ文書部品として出力するか或いは論理構造を有する1つの文書として出力する文書化手段とを具備している。

【0021】第2の発明の文書処理装置は、第1の発明において、前記指定手段は、前記分割情報に従って分割される複数の部分構造をそれぞれ段落構造とする旨を指定すると共に、当該複数の段落構造に対して適用されるべく段落属性を指定し、前記処理手段は、前記分割情報に従って分割した複数の段落構造の内容に対して前記段落属性を適用させると共に、当該各段落構造それぞれに前記構造情報を付加することを特徴としている。

【0022】第3の発明の文書処理装置は、第1の発明において、前記指定手段は、前記分割情報に従って分割される複数の部分構造をそれぞれ段落構造として指定すると共に、当該複数の段落構造に対して適用されるべく予め設定された段落属性を表しているひな型情報を指定し、前記処理手段は、前記分割情報に従って分割した複数の段落構造の内容に対して前記ひな型情報に対応する段落属性を適用させると共に、当該各段落構造それぞれに前記構造情報を付加することを特徴としている。

【0023】第4の発明の文書処理装置は、文書内容が構造化されていない複数の文書を保持する保持手段と、前記複数の文書毎に複数の部分構造に分割するための情報を示す分割情報を指定すると共に、前記複数の文書間の階層的な関係を定義するための情報を示すレベル情報を指定する指定手段と、前記文書毎に該当する文書を前

記分割情報に基づいて複数の部分構造に分割すると共に、当該複数の文書の分割された複数の部分構造それぞれに、前記レベル情報と各文書の論理的なレイアウト上の位置に対応する構造情報を付加する処理手段と、予め設定された指示に応じて、前記処理手段によって構造情報を付加された複数の部分構造を、それぞれ文書部品として出力するか或いは論理構造を有する1つの文書として出力する文書化手段とを具備したことを特徴としている。

【0024】第5の発明の文書処理装置は、第4の発明において、前記指定手段は、前記分割情報に従って分割される複数の部分構造をそれぞれ段落構造として指定すると共に、当該複数の段落構造に対して適用されるべく段落属性を指定し、前記処理手段は、前記分割情報に従って分割した複数の段落構造の内容に対して前記段落属性を適用させると共に、当該各段落構造それぞれに前記レベル情報及び構造情報を付加することを特徴としている。

【0025】第6の発明の文書処理装置は、第4の発明において、前記指定手段は、前記分割情報に従って分割される複数の部分構造をそれぞれ段落構造として指定するとともに、当該複数の段落構造に対して適用されるべく予め設定された段落属性のひな型情報を指定し、前記処理手段は、前記分割情報に従って分割した複数の段落構造の内容に対して前記ひな型情報に対応する段落属性を適用させると共に、当該各段落構造それぞれに前記レベル情報及び構造情報を付加することを特徴としている。

【0026】

【作用】この発明においては、指定手段によって、文書内容が構造化されていない文書を複数の部分構造に分割するための情報を示す分割情報が指定されると、処理手段は、前記文書を前記分割情報に基づいて複数の部分構造に分割すると共に、当該各構造それぞれに、少なくとも、前記文書における論理的なレイアウト上の位置に対応する構造情報を付加し、そして文書化手段は、予め設定された指示に従って、構造情報が付加された複数の部分構造をそれぞれ文書部品として出力するか或いは論理構造を有する1つの文書として出力する。これにより、構造化されていない文書から、論理構造を有する1つの構造化文書あるいは複数の文書部品を得ることができ

【0027】また上記発明においては、指定手段によって、上記分割情報に加えて、複数の部分構造をそれぞれ段落構造とする旨及び当該各段落構造に対する段落属性が指定されると、処理手段は、上記分割情報に従って分割した複数の段落構造の内容に対して前記段落属性を適用させると共に、当該各段落構造それぞれに上記構造情報を付加する。これにより、構造化されていない文書から構造化文書を得ることができ、しかも該構造化文書を構

成する段落構造に、任意に設定した段落属性を適用させることができる。

【0028】さらに上記発明においては、指定手段によって、上記分割情報に加えて、複数の部分構造をそれぞれ段落構造とする旨及び該各段落構造に対するひな型情報が指定されると、処理手段は、上記分割情報に従って分割した複数の段落構造の内容に対して前記ひな型情報に対応する段落属性を適用させると共に、当該各段落構造それぞれに上記構造情報を付加する。これにより、構造化されていない文書から構造化文書を得ることができ、しかも該構造化文書を構成する段落構造に、予め設定された段落属性（ひな型）を適用させることができる。

【0029】また他の発明においては、指定手段によって、文書内容が構造化されていない複数の文書を複数の部分構造に分割するための情報を示す分割情報、及び前記複数の文書間の階層的な関係を定義するための情報を示すレベル情報が指定されると、処理手段は、前記文書毎に該当する文書を前記分割情報に基づいて複数の部分構造に分割すると共に、当該複数の文書の分割された複数の部分構造それぞれに、前記レベル情報と各文書の論理的なレイアウト上の位置に対応する構造情報を付加し、そして文書化手段は、予め設定された指示に従って、構造情報が付加された複数の部分構造をそれぞれ文書部品として出力するかいは論理構造を有する1つの文書として出力する。これにより、構造化されていない複数の文書から、論理構造を有する1つの構造化文書あるいは複数の文書部品を得ることができる。

【0030】また上記他の発明においては、指定手段によって、上記分割情報に加えて、複数の部分構造をそれぞれ段落構造とする旨及び該各段落構造に対する段落属性が指定されると、処理手段は、上記分割情報に従って分割した複数の段落構造の内容に対して前記段落属性を適用させると共に、当該各段落構造それぞれに上記レベル情報及び構造情報を付加する。これにより、構造化されていない複数の文書から1つの構造化文書を得ることができ、しかも該構造化文書を構成する段落構造に、任意に設定した段落属性を適用させることができる。

【0031】さらに上記他の発明においては、指定手段によって、上記分割情報に加えて、複数の部分構造をそれぞれ段落構造とする旨及び該各段落構造に対するひな型情報が指定されると、処理手段は、上記分割情報に従って分割した複数の段落構造の内容に対して前記ひな型情報に対応する段落属性を適用させると共に、当該各段落構造それぞれに上記レベル情報及び構造情報を付加する。これにより、構造化されていない複数の文書から1つの構造化文書を得ることができ、しかも該構造化文書を構成する段落構造に、予め設定された段落属性（ひな型）を適用させることができる。

【0032】

【実施例】以下、本発明の実施例を添付図面を参照して説明する。

【0033】図1は本発明に係る文書処理装置の第1の実施例を機能ブロック図で示したものである。同図において、内容情報入力手段10は内容情報となるデータを入力し、内容情報解釈手段20はそのデータを解釈する。

【0034】論理構造定義手段30は生成すべく論理構造の定義を入力する。すなわち論理構造を有しない文書を複数の部分構造に分割するための情報を示す分割情報を指定するためのものである。また複数の文書毎に複数の部分構造に分割するための情報を示す分割情報を指定すると共に、当該複数の文書間の階層的な関係を定義するための情報を示すレベル情報を指定するためのものである。論理構造定義解釈手段40は論理構造定義手段30によって定義された定義内容を解釈する。

【0035】出力先定義手段50は文書の出力先を定義し、出力先定義解釈手段60はその出力先を解釈する。

【0036】変換手段70は、内容情報解釈手段20の解釈結果に基づいて内容情報保持手段80から内容情報を読み込むと共に、この内容情報を、論理構造定義解釈手段40の解釈結果に従って文書部品に変換する。すなわち、文書を上記分割情報に基づいて複数の部分構造に分割し、当該各構造それぞれに、少なくとも、文書における論理的なレイアウト上の位置に対応する構造情報を付加して、文書部品に変換する。また文書毎に該当する文書を上記分割情報に基づいて複数の部分構造に分割すると共に、当該複数の文書の分割された複数の部分構造それぞれに、上記レベル情報と各文書の論理的なレイアウト上の位置に対応する構造情報を付加して文書部品に変換する。

【0037】文書化手段90は、変換手段70の変換結果を、予め設定された指示に応じて、前記処理手段によって構造情報を付加された複数の部分構造をそれぞれ文書部品として出力するか、あるいは当該複数の部分構造を1つの構造化文書として、出力先定義解釈手段60の解釈結果である出力先へ出力する。

【0038】なお、この実施例においては、上述した保持手段は内容情報保持手段80に対応し、上述した指定手段は内容情報入力手段10及び論理構造定義手段30に対応し、上述した処理手段は変換手段70に対応し、上述した文書化手段は文書化手段90に対応する。

【0039】図2は上記実施例の装置を実現するためのハードウェア構成図を示したものであり、この装置は、例えばワークステーションやコンピュータなどの処理装置に適用される。同図において、装置は、入力装置210、主メモリ220、外部記憶装置230、ディスプレイ240、中央処理装置（CPU）250を有して構成されている。なおこれらの構成要素は、バス260を介してそれぞれ接続されており、CPU250はバス260

0を介して上記各構成要素を制御する。

【0040】ここで、図1に機能ブロック図の構成要素と図2に示したハードウェア構成図の構成要素との対応関係について説明する。

【0041】図1に示した内容情報入力手段10、論理構造定義手段30及び出力先定義手段50は図2に示した入力装置210（実際には、ユーザ・インタフェースとしてのシートと、入力装置210とから構成される。）に対応し、図1に示した内容情報保持手段80は図2に示した外部記憶装置230及び主メモリ220に対応し、図1に示した内容情報解釈手段20、論理構造定義解釈手段40、出力先定義解釈手段60、変換手段70及び文書化手段90は図2に示したCPU250（実際には、CPU250が各手段の機能を遂行させるためのソフトウェアを実行することにより実現される。）に対応する。

【0042】この発明においては、通常のテキストからなる情報を、構造化文書の要素となる様な、構造を持つテキストに変換するようにしたものである。

【0043】構造化文書とは、通常の内部テキスト情報以外に、構造情報や属性情報も有し、構造関係がグラフや木により表現されるものである。文書は文書部品から構成され、この文書部品は、構造上の位置を現す情報、付帯する属性情報、内容情報から構成されている。

【0044】ここで、構造化文書モデルの具体例を以下に示す。

【0045】〈と〉とに囲まれた部分は1つの要素であり、このような要素の複合で文書が形成されるものとする。

【0046】構造化文書は〈構造単位開始記号〉、〈構造単位終端記号〉で囲まれた部分が1つの構造の構成要素となる。簡単のため、〈構造単位開始記号〉を「〈〈構造単位終端記号〉を」と記述する。この文書の構造構成要素の組み合わせにより、文書または文書部品が作成される。

【0047】「〈レベル識別子〉〈要素属性〉〈内容情報〉」が通常の構造化文書中の、1構成要素を表す例である。なお〈内容情報〉は省略される場合もある。

【0048】〈レベル識別子〉は、その要素がどのレベルにあるかを示すもので、2個の数値からなる。例えば、第1章第3節であれば、〈2, 3〉である。最初の2が深さが2であることを示し、次の3がその深さでの3番目の順位を持つ構成要素であることを意味する。このレベル識別子は全ての要素に付加される。

【0049】〈要素属性〉は、その要素が何かを示す。値としては、見出し構造、段落構造、枠構造、見出しタイトル内容部、文字内容部、図形内容部といった枠の内容部、などがある。

【0050】〈内容情報〉は、属性要素が見出しタイトル内容部や文字内容部のように、内容部としての情報が

存在するときのみ出現する。文字内容部には、段落全体にかかる文字属性が記述できる。

【0051】そこで、

「〈1, 1〉〈見出し構造〉」

「〈2, 1〉〈タイトル段落構造〉」

「〈3, 1〉〈見出しタイトル内容部〉〈はじめに〉」

「〈2, 2〉〈段落構造〉」

「〈3, 2〉〈文字内容部〉〈概要の説明です〉」

「〈2, 3〉〈段落構造〉」

「〈3, 3〉〈文字内容部〉〈文字種：明朝。；〉それではこれで終わります」

のように記述される構造化文書は、

はじめに

(章の先頭)

概要の説明です

(この章の第1節)

それではこれで終わります

(この章の第2節)

という構成であることが分かる。

【0052】このような構造化文書を木構造で表現すると、図3に示すような論理構造で表現される。図3においては、構成要素310は深さが1、構成要素320、340、360は深さが2、構成要素330、350、370は深さが3であり、構成要素310、320、330が1番目の順位、構成要素が340、350が2番目の順位、構成要素が360、370が3番目の順位である。従って、例えば上述した「〈3, 2〉〈文字内容部〉〈概要の説明です〉」は深さが3で2番目の順位の構成要素を意味しているので、これは、図3に示される例の構成要素350と一致している。他の構成要素についても同様なことが言える。

【0053】また、この様な木構造で表現される論理構造においては、探索深さ優先の探索法により探索されるので、構成要素310～370の順に選択（探索）されることとなる。

【0054】図4はグラフィカル・ユーザ・インタフェース（GUI）の機能を果たすシート（以下、GUIシートという）を示している。図4において、GUIシート400は、操作領域410と、入力情報設定領域420と、出力情報設定領域430とから構成されている。

【0055】操作領域410は、設定された情報に従い処理の実行を指示する「実行」項目と、設定された情報の取消しを指示する「取消」項目とを有する領域411と、「次の指定」項目と「前の指定」項目と「追加設定」項目とを有する領域412とを有している。「次の指定」項目を指示すると、既に設定済みの次のシート内容（次画面）が表示され、「前の指定」項目が指示されると、既に設定済みの前のシート内容（前画面）が表示され、「追加設定」項目が指示されると、新規のシートの内容を追加設定することが出来る。なお「追加設定」項目を指示しないときは、1つのテキストファイル内の内容情報についての入力情報の設定のみとなり、「追加設定」項目を1回指示する毎に、1つのテキストファイ

ル内の内容情報についての入力情報を新たに設定可能となる。従って「追加設定」項目が5回指示されたということは、6つのテキストファイル内の内容情報についての入力情報を設定していることを意味している。

【0056】入力情報設定領域420は、「入力」欄421と「生成構造」欄422と「段落生成」欄423とを有して構成されている。

【0057】「入力」欄421は「ファイル」か「テキスト」かを指定するためのものである。この図4に示す例においては、ファイルが指定されているので、ファイル名を入力するための欄が表示されている。この様に符号421Aで示される「テキスト」「ファイル」のように、並列に複数並んでいるものはチョイスと呼ばれる。なお「テキストが指示された場合は、文字列を入力するための欄が表示されるようになっている。この様に符号421Bで示される「ファイル名」を入力するための欄はテキスト入力パーツと呼ばれる。この例では、「data1」という文字列がサブパラメータとして記述されている。

【0058】「生成構造」欄422は、「段落」か「見出し」かを指定するためのものである。この図に示す例においては、「見出し」が指定されているので、レベル及びタイトルを入力するためのテキスト入力パーツが表示され、それぞれサブパラメータが記述されている。このテキスト入力パーツにおけるレベルに設定される値がレベル情報となる。一方、「段落」が指定された場合はテキスト入力パーツは表示されない。

「  
ああああああああ。○  
○  
いいいいいいいい。○

」

…テキストファイル1

ここで、「」内の内容は1つのテキストファイル1内の内容情報

○は改行記号

なおこの場合、テキストファイル1の内容情報中の先頭※

「  
ああああああああ。○  
○  
いいいいいいいい。○

」

…段落構造1

のように1つの段落構造が得られることとなる。

【0064】また「空行で改段落」が指定された場合

「  
ああああああああ。○  
」

…段落構造2

と

「  
いいいいいいいい。○  
」

…段落構造3

\*【0059】ここでレベル情報の具体例を説明する。文書構造であれば、単純な本文内の段落、章節といった見出しがつく構造、が考えられる。章節であるならば、その深さ(1、1、1、1、1、1、…)がある。従ってこの実施例においては、レベル情報は、本文段落或いは見出しの構造の設定内容と、見出しの場合は、その深さ及び見出し部分に付加する文字列(タイトル)の指定内容とから構成されている。またタイトルに設定した文字列は、直接入力するようにしているが、内容情報の指定と同様に、外部のファイルを指定し、その内容をタイトルに設定する様にしても良い。

【0060】「段落生成」欄423は、「1段落」か「空行で改段落」か「改行で改段落」かを指定するためのものである。すなわち、1つの内容情報を複数の部分構造に分割するための分割情報を指定するためのものである。この図4に示す例では、「1段落」が指定されている。

【0061】「1段落」は文書の内容情報を1つの段落構造とする旨を指定するためのものであり、「空行で改段落」は改行記号と改行記号との間の内容情報及び改行記号と改行記号との間の空白をそれぞれ1つの段落構造とする旨を指定するためのものであり、「改行で改段落」は改行記号と改行記号との間の内容情報を1つの段落構造とする旨を指定するためのものである。

【0062】例えば、以下のようなテキストファイルが指定されたとする。

※の文字の前に、改行記号が存在しているものと見なされる。

【0063】上記内容情報においては、「1段落」が指定された場合は、



13

14

の2つの段落構造（論理構造における出現順序は段落構造2、段落構造3となる）が得られることとなる。

\*【0065】更に「改行で改段落」が指定された場合は、

「  
ああああああああ。○  
」 …段落構造4

と

「  
(空白)  
」 …段落構造5

と

「  
いいいいいいいい。○  
」 …段落構造6

の3つの段落構造（論理構造における出現順序は段落構造4、段落構造5、段落構造6となる）が得られることとなる。

【0066】このように、第1の実施例においては、一般的なテキストにみられる改行に着目して、1つの内容情報を複数の部分構造に分割するようにしている。なお第1の実施例では、分割情報として上述した「1段落」、

※終わる（次に改行がくる）ようにした場合、その位置で内容情報を分割すれば良い。またGUI400（図4参照）の「段落生成」欄423のチョイスの内容として、上記3つの項目に加えて、「特殊文字又は文字列」の項目を追加しておく。この場合、上記3つの項目と、上記「特殊文字又は文字列」項目のいずれかを分割情報として指定することになる。この様な分割情報が選択された場合に、変換手段70は、特殊文字又は文字列を改行記号と見なして、上記「1段落」、「空行で改段落」、「改行で改段落」の項目が選択された場合の処理と同様の処理を行う。なおこの場合、文書中に挿入されている特殊文字又は文字列は、段落構造に分割された際に削除されるようになっている。

【0067】すなわち、作成中又は既存の非構造化文書における、段落構造とすべく内容情報の位置（分割したい位置）に、#や干など予め設定された特殊文字を挿入するか、あるいは「段落区切」など、内容情報を分割するということを明示的に表現している予め設定された文字列を挿入する。そして、例えば、特殊文字や文字列で※

【0068】例えば、以下のようなテキストファイルが指定されたとする。

「  
ああああああああ。#○  
#○  
いいいいいいいい。○  
うううううううう。#○  
」 …テキストファイル2

ここで、『』内の内容は1つのテキストファイル2内の内容情報

の文字の前に、特殊文字が存在しているものと見なされる。

○は改行記号、#は特殊記号

【0069】上記内容情報においては、「特殊文字又は文字列」が指定された場合は、

なおこの場合、テキストファイル2の内容情報中の先頭

「  
ああああああああ。○  
」 …段落構造7

と

「  
(空白)  
」 …段落構造8

と

「  
いいいいいいいい。○  
うううううううう。○  
」 …段落構造9

の3つの段落構造(論理構造における出現順序は段落構造7、段落構造8、段落構造9となる)が得られることとなる。

【0070】出力領域設定領域430は、「生成形式」欄431と「生成先」欄432とを有して構成されている。

【0071】「生成形式」欄431は、「文書」か「文書部品」かを指定するためのものである。この図4に示す例では、「文書部品」が指定されている。

【0072】「生成先」欄432は、「標準出力」か「ファイル」かを指定するためのものである。この図4に示す例では、「ファイル」が指定されているので、ファイル名を指定するためのテキスト入力パーツが表示され、そのサブパラメータが記述されている。「標準出力」が指定された場合はテキスト入力パーツは表示されない。

【0073】このような図4に示す状態から、今度は、「入力」欄421においては「テキスト」項目を指定し、「段落生成」欄423においては「段落」項目を指定した場合のGUIシート400は、図5に示すような内容に変更される。この図5において、「入力」欄421においては、「テキスト」項目に対応するテキスト入力パーツが表示され、その後、そのサブパラメータが記述された様子が示されており、また「段落生成」欄423においては「見出し」項目に対応するテキスト入力パーツが消去された様子が示されている。

【0074】なお、図4及び図5において、「入力」欄421は内容入力手段10の機能を有し、「生成構造」欄422、「段落生成」欄423及び「生成形式」欄431は論理構造定義手段30の機能を有し、「生成先」欄432は出力先定義手段50の機能を有している。

【0075】以上説明した様にこの実施例においては、入力情報をGUIで設定することにより操作性を向上させている。勿論、これに限定されることはなく、GUIを持たないコマンド(アプリケーション)として実現されていても良い。この場合、必要な入力情報は上記コマンドのオプションとして引数で指定する。

【0076】図6は、内容情報入力手段10、論理構造定義手段30、出力先定義手段50によって指定された情報を保持するメモリ600の保持形態を示している。

【0077】メモリ600は、記憶領域610とワーク領域620とを有して構成されている。記憶領域610は、入力情報領域611と、「内容情報の個数」が値として設定される個数領域612と、「文書」か「文書部品」かのいずれかが値として設定される出力形式領域613と、「出力として作成するファイル名」が値として設定される出力先指定領域614とを有して構成されている。

【0078】入力情報領域611においては、「外部ファイル指定」か或いは「テキスト指定」のいずれかが値

として設定される入力フラグ領域611aと、「外部ファイル指定の場合にファイル名」が、一方、「テキスト指定の場合は文字列」が値として設定される入力用文字列領域611bと、「段落構造指定」か或いは「見出し構造指定」のいずれかが値として設定される構造情報フラグ領域611cと、「深さの値(数値)」が値として設定される構造レベル領域611dと、「文字列」が設定されるタイトル用文字列領域611eと、「1段落」か「空行で改段落」か「改行で改段落」かのいずれかが値として設定される。

【0079】ワーク領域620は、構造のレベルが設定される構造レベル領域621と、レベル識別子が設定されるレベル識別子領域622と、生成データ一時記憶アドレス領域623と、生成データ一時記憶領域623とを有して構成されている。このワーク領域620においては、構造レベル領域611dに指定されたレベルだけのワーク領域が確保される。設定個数領域612に指定された個数のアドレス記憶領域が確保される。処理の過程で動的にアドレスが記入される。

【0080】図7は、ワーク領域620の構造レベル領域621、レベル識別子領域622に動的に設定される内容の設定処理を説明するための図を示している。

【0081】図7(a)においては、構造レベル領域621及びレベル識別子領域622の内部を示したものであり、内容を現在保有している各レベルでのレベル識別子の最大の個数maxが設定される領域Mと、現在着目しているレベルを示す値indexが設定される領域Iとによって構造レベル領域621が形成され、レベルの値が設定され領域Vによりレベル識別子領域622が形成されている。

【0082】ここで、レベルを1段深くする場合、領域Iのindexを1増やす。このとき領域Mのmaxより大きくなれば、その個数分だけ、領域Vの領域の再確保が行われる。また1つの見出し構造に複数の段落がある場合など、indexに関わらず、構造レベル領域621のindex番目の値を順次増やしていくことで対応をとるようになっている。

【0083】図7(b)は、見出し構造の論理構造の指定の順序を説明するための図を示したものである。この図においては、構成要素710の下位に構成要素720、730、740が接続されていることが示されている。そして、これらの構成要素を「A」、「B」、「C」、「D」の順に指定するものとし、「A」は構成要素710を、「B」は構成要素720を、「C」は構成要素730を、「D」は構成要素730をそれぞれ指定していることが示されている。

【0084】なお構成要素710はレベル識別子<1>、構成要素720、730、740はそれぞれレベル識別子<2.1>、<2.2>、<2.3>である。

【0085】図7(c)は、「A」、「B」、「C」、

「D」の順に指定した場合の構造レベル領域621及びレベル識別子領域622の内部状態を示したものである。

【0086】次に各指定時の内部状態について説明する。

【0087】◇初期時

領域Mの値=1、領域Iの値(深さ)=1、領域Vの値(構成要素の兄弟)=0になっている。この状態では、構成要素についての情報が書き込まれていないので、領域Vの値=0になっている。

【0088】◇A指定時(構成要素710の指定)

領域Mの値=1、領域Iの値=1、領域Vの値=「1」となる。この内部状態は、領域I及び領域Vの値より、レベル識別子<1・1>を意味している。これは上記構成要素710のレベル識別子と一致する。

【0089】◇B指定時(構成要素720の指定)

A指定終了時の内部状態においては、領域Vに、構成要素720についての情報を書き込む領域がないので、もう1つの構成要素についての情報を書き込むための領域が追加される。また深さが「2」になるので、領域Iの値=2となる。従って、領域Mの値=1、領域Iの値=2、領域Vの値=「1」、「1」となる。この内部状態は、領域I及び領域Vの値より、レベル識別子<2・1>を意味している。これは上記構成要素720のレベル識別子と一致する。

【0090】◇C指定時(構成要素730の指定)

今度は構成要素730についての情報が書き込まれることになり、領域Vの値=「1」、「1」が「1」、「2」に変更される。従って内部状態は、領域Mの値=2、領域Iの値=2、領域Vの値=「1」、「2」となる。この内部状態は、領域I及び領域Vの値より、レベル識別子<2・2>を意味している。これは上記構成要素730のレベル識別子と一致する。

【0091】◇D指定時(構成要素740の指定)

C指定時と同様の処理が行われるので、内部状態は、領域Mの値=2、領域Iの値=2、領域Vの値=「1」、「3」となる。この内部状態は、領域I及び領域Vの値より、レベル識別子<2・3>を意味している。これは上記構成要素740のレベル識別子と一致する。

【0092】このような処理は、変換手段70によって実施される。

【0093】次に、この実施例の概要処理について、図8に示すフローチャートを参照して説明する。

【0094】内容情報入力手段10によって内容情報が入力され(ステップ810)、論理構造定義手段30によって論理構造の定義が入力され(ステップ820)、更に出力先定義手段50によって出力先の定義が入力されると、内容情報解釈手段20は上記入力された内容情報の解釈処理を実行し(ステップ830)、論理構造定義解釈手段40は上記入力された論理構造の定義内容の

解釈を実行し(ステップ840)、更に出力先定義解釈手段60は上記出力先の定義内容を解釈する。このような解釈の際に、各解釈手段は、入力情報の解釈に不具合があるか否かを判断し(ステップ850)、不具合がある場合は上記ステップ810に戻る。不具合がある場合はその旨がユーザに通知されるので、ユーザによって正規の入力情報が入力されることになる。

【0095】ステップ850において不具合がない場合は、変換手段70は、内容情報解釈手段20及び論理構造定義解釈手段40の解釈結果に基づいて、内容情報保持手段80から内容情報を読み込んで、その内容情報を解釈すると共に(ステップ860)、この解釈した内容情報を文書又は文書部品に変換する変換処理を実行する(ステップ870)。

【0096】このようにして変換処理が終了すると、文書化手段90は、変換手段70の変換結果を文書化すると共に、該文書化された内容を出力先定義解釈手段60の解釈結果である出力先へ出力する文書化処理を実行する(ステップ880)。

【0097】次に、入力情報の解釈処理について、図9に示すフローチャートを参照して説明する。

【0098】図示しない制御手段は、入力情報の個数をメモリ600(図6参照)の設定個数領域612に設定し(ステップ901)、その値分だけ入力情報領域611を確保する(ステップ902)。

【0099】次に $i=1$ とし(ステップ903)、設定個数領域612に設定された個数だけ後述する処理を実行しループしたか否かを判断する(ステップ904)。

【0100】ステップ904においてループを終了していない場合は、内容情報情報解釈手段20及び論理構造定義解釈手段40によって、1番目の入力情報領域に、GUIシート400(図4参照)の入力情報設定領域420に設定された入力情報を設定する処理が実行され(ステップ905)、この実行結果がエラーか否かを判断する(ステップ906)。

【0101】ステップ906において、エラーでない場合は、 $i=i+1$ として(ステップ907)、上記ステップ904に戻り、上記設定個数分のループが終了するまでステップ904以降を実行する。一方、エラーの場合は、シート入力へ戻る。

【0102】ところで、上記ステップ904においてループを終了した場合、論理構造定義解釈手段40は、メモリ600の出力形式領域613に、GUIシート400の「生成形式」欄431に設定された出力形式を設定し(ステップ908)、また出力先定義解釈手段60は、出力先指定領域614に、出力情報設定領域430の「生成先」欄432に設定された出力先を設定する(ステップ909)。この出力先の設定に際してはファイルのときはファイル名を、一方、標準出力のときは「0」を設定する。

【0103】次に、上記ステップ905の処理について、図10に示すフローチャートを参照して説明する。

【0104】最初に、図10に示すように、内容情報解釈手段20は、GUIシート400の「入力」欄421の設定内容に基づいて、内容情報は外部ファイルか否かを判断する(ステップ1001)。

【0105】ステップ1001において外部ファイルの場合には、メモリ600の入力フラグ領域611aに、“外部ファイル指定”のフラグを設定すると共に(ステップ1002)、入力用文字列領域611bに、「入力」欄421のテキスト入力パーツに設定されているサブパラメータ(すなわちファイル名)を設定する(ステップ1003)。

【0106】一方、ステップ1001において外部ファイルでない場合は、入力フラグ領域611aに、テキスト指定のフラグを設定すると共に(ステップ1004)、入力用文字列領域611bに、「入力」欄421のテキスト入力パーツに設定されているサブパラメータ(すなわち文字列)を設定する(ステップ1005)。

【0107】ステップ1003あるいはステップ1005を終了した場合と、論理構造定義解釈手段40は、GUIシート400の「生成構造」欄422の設定内容に基づいて、生成構造は段落か否かを判断する(ステップ1006)。

【0108】ステップ1006において段落の場合には、メモリ600の構造情報フラグ領域611cに、“段落構造指定”のフラグを設定し(ステップ1007)、その後、段落に関する処理を実行する(ステップ1008)。

【0109】一方、ステップ1006において段落でない場合は、構造情報フラグ領域611cに、“見出し構造指定”のフラグを設定し(ステップ1009)、「生成構造」欄422のテキスト入力パーツのサブパラメータ(図4参照)に基づいて、構造レベル指定は有効か否かを判断する(ステップ1010)。

【0110】ステップ1010において有効の場合は、構造レベル領域611dに、「生成構造」欄422のテキスト入力パーツにおける「レベル」に設定されたサブパラメータ(この場合はサブパラメータ=数値)を設定すると共に(ステップ1011)、上記テキスト入力パーツにおける「タイトル」のサブパラメータに基づいて、見出し文字指定があるか否かを判断する(ステップ1012)。

【0111】ステップ1012において、見出し指定がある場合には、タイトル用文字列領域611eに、上記テキスト入力パーツにおける「タイトル」に設定された文字列を設定し(ステップ1013)、一方、見出し指定が無い場合は、タイトル用文字列領域611eに、「0」文字列無しを設定する(ステップ1014)。

【0112】そしてステップ1013あるいはステップ

1014を終了した場合はステップ1008に進む。

【0113】なおステップ1010において構造レベル指定が無効の場合はエラー処理が実行される。

【0114】次に、上記ステップ1008の処理について、図11に示すフローチャートを参照して説明する。

【0115】図11に示すように、論理構造定義解釈手段40は、メモリ600の生成段落指定フラグ領域611fに、GUIシート400の「段落生成」欄423のチョイス中で選択された段落生成の種類(例えば“1段落”や改行で改段落など)を設定する(ステップ1100)。

【0116】次に、変換手段70による変換処理について、図12及び図13に示すフローチャートを参照して説明する。

【0117】最初に、図12に示すように、変換手段70は、ワーク領域620の構造レベル領域621及びレベル識別子領域622を初期化すると共に、記憶領域610の設定個数領域612に指定された値分の生成データ一時記憶アドレス領域を確保し(ステップ1201)、更に、設定個数領域621に指定された値分の回数をループしたか否かを判断する(ステップ1202)。

【0118】ステップ1202において、上記回数分のループを終了した場合には処理を終了し、一方、上記回数分のループを終了していない場合は、ループ回数番目の生成データ一時記憶アドレス領域に、これからデータを格納する生成データ一時記憶領域の中のアドレスを設定する(ステップ1203)。

【0119】ステップ1203を終了したら、記憶領域610の構造情報フラグ領域611cの値に基づいて、生成構造は“見出し”か否かを判断し(ステップ1204)、“見出し”でない場合には、属性=構造構造として、文書構造構成要素の生成処理を実行し(ステップ1205)、一方、“見出し”の場合は、属性=見出し構造として、文書構造構成要素の生成処理を実行する(ステップ1206)。

【0120】ステップ1205、ステップ1206を終了した場合は、構造レベル領域621の構造レベルの値に「+1」を加算し(これは、レベルを1段深くすることを意味する)、この構造レベルに対応するレベル識別子の値を初期化する(ステップ1207)。

【0121】次に生成構造は“見出し”か否かを判断し(ステップ1208)、“見出し”の場合は、属性=見出しタイトルとして、文書構造構成要素の生成処理を実行し(ステップ1209)、その後、構造レベル領域621の構造レベルの値に「+1」を加算し、この構造レベルに対応するレベル識別子の値を初期化する(ステップ1210)。このステップ1210を終了すると、図13に示される処理が行われる。

【0122】すなわち、図13に示す様に、変換手段7

0は、データ=タイトル用文字列領域611eに設定されている文字列として、内容情報（すなわちタイトル文字内容）の生成処理を実行し（ステップ1211）、構造レベル領域621の構造レベルの値に「-1」を加算（これは、レベルを1段浅くすることを意味する）と共に（ステップ1212）、この構造レベルに対応するレベル識別子の値に「+1」を加算する（ステップ1213）。

【0123】この処理が終了すると、属性=段落構造として、文書構造構成要素の生成処理を実行し（ステップ1214）、その後、構造レベル領域621の構造レベルの値に「+1」を加算する共に、この構造レベルに対応するレベル識別子の値を初期化する（ステップ1215）。

【0124】そして入力フラグ領域611aに設定された値に基づいて、内容指定はファイルか否かを判断し（ステップ1216）、ファイル指定の場合は、入力用文字列領域611bに設定された値すなわちファイル名に対応するファイルはオープンできたか否かを判断する（ステップ1217）。

【0125】ここでオープンできた場合は、データ取得先はオープンされたファイルであると認識し（ステップ1218）、データ=上記ファイルからのデータを内容情報として、内容情報の生成処理を実行する（ステップ1219）。なおステップ1217においてオープンできなかった場合はエラー処理が行われる。

【0126】上記ステップ1216において、内容指定がファイルでない場合は、データ取得先はメモリ600であると認識し（ステップ1220）、データ入力用文字列領域611bに設定された文字列として、内容情報の生成処理を実行する（ステップ1221）。

【0127】ステップ1219あるいはステップ1221を終了すると、図12に示した処理手順のステップ1202に戻る。

【0128】なお図12に示したステップ1208において、生成構造が“見出し”でない場合すなわち“段落構造”の場合は、図13に示した処理手順のステップ1216に進む。

【0129】次に、変換手段70による文書構造生成要素の出力処理について、図14に示すフローチャートを参照して説明する。

【0130】最初に、構造単位開始記号を生成データ一時記憶領域624に出力すると共に（ステップ1401）、ワーク領域620の構造レベル領域621及びレベル識別子領域622からレベル情報を生成データ一時記憶領域624に出力し（ステップ1402）、更にパラメータによって指定されている属性指定（すなわち段落構造あるいは見出し構造）を、生成データ一時記憶領域624に出力し（ステップ1403）、その後、構造単位終了記号を生成データ一時記憶領域624に出力す

る（ステップ1404）。

【0131】この図に示す処理手順によって、図12に示したステップ1205、図13に示したステップ1214においては“段落構造”が生成され、また図12に示したステップ1206においては“見出し構造”が生成され、更に図12に示したステップ1209においては“見出しタイトル”が生成される。

【0132】次に、変換手段70による内容情報の出力処理について、図15及び図16に示すフローチャートを参照して説明する。

【0133】最初に図15に示すように、変換手段70は、構造単位開始記号を生成データ一時記憶領域624に出力し（ステップ1501）、また構造レベル領域621、レベル識別子領域622からレベル情報を生成データ一時記憶領域624に出力し（ステップ1502）、構造属性として文字内容部を生成データ一時記憶領域624に出力する（ステップ1503）。

【0134】そしてデータが指定されたメモリあるいはファイルに存在するか否かを判断し（ステップ1504）、存在しない場合には構造単位終了記号を出力し（ステップ1505）、一方、データが存在する場合は、指定されたデータ入力先（メモリあるいはファイル）から1文字を取得し（ステップ1506）、そのデータは改行か否かを判断する（ステップ1507）。

【0135】ステップ1507において改行の場合は、生成段落指定フラグ領域611fに指定されている生成段落は1段落かを判断する（ステップ1508）。

【0136】ステップ1507において改行でない場合、ステップ1508において生成段落が1段落の場合は、データを生成データ一時記憶領域624へ出力し（ステップ1509）、その後、ステップ1504に戻る。

【0137】ステップ1508において生成段落が1段落でない場合は、生成段落の指定は空行か否かを判断し（ステップ1510）、空行の場合には、次の文字が改行なら連続する改行を読み捨てて（ステップ1511）、構造単位終了記号を出力する（ステップ1512）。ステップ1510において空行でない場合つまり改行の場合はステップ1512へ進む。

【0138】このようにして構造単位終了記号を終了すると、構造レベル領域621の構造レベルの値に「-2」を加算し（これは構造レベルを2段階戻すことを意味している）（ステップ1513）、該構造レベルに対応するレベル識別子の値に「+1」を加算し（ステップ1514）、更に、属性=段落情報として文書構造構成要素の生成処理を実行する（ステップ1515）。

【0139】そして、構造レベル領域621の構造レベルの値に「+1」を加算し、対応するレベル識別子の値を初期化し（ステップ1516）、構造単位開始記号を生成データ一時記憶領域624に出力すると共に（ステ

ップ1517)、構造レベル領域621、レベル識別子領域よりレベル情報を出し(ステップ1518)、構造属性として文字内容部を生成データ一時記憶領域624に出力する(ステップ1519)。

【0140】ステップ1519を終了すると、図15に示したステップ1504に戻る。

【0141】次に、文書化手段90による出力処理について、図17に示すフローチャートを参照して説明する。

【0142】文書化手段90は、出力先指定領域614より出力ファイル名を取得し(ステップ1701)、ファイル指定有無かを判断し(ステップ1702)、ファイル指定の場合は、そのファイルをオープンできたかを判断する(ステップ1703)。

【0143】ステップ1703において、オープンできない場合にはエラー処理を実行し、一方、オープンできた場合は、出力先はそのオープンされたファイルであると認識し(ステップ1704)、その後、文書の生成可否かを判断する(ステップ1705)。

『あああ… …あ。○

…あああ。○ 』 …テキストファイル10

と

『いいい… …い。○

…いいい。○ 』 …テキストファイル20

とである。

【0150】なお、『』内は1つのテキストファイル内の内容情報

○は改行記号

ここで、「生成構造」欄422の設定内容は、テキストファイル10の内容情報に対するレベル情報として、レベル「1」、タイトル「第1章」、またテキストファイル20の内容情報に対するレベル情報として、レベル「1」「第2章」が設定されているものとする。さらに「段落生成」欄423の設定内容は「1段落」が設定されているものとする。

【0151】このような設定内容においては、上述したテキストファイル10、20の内容情報は以下のように変換される。すなわち

「第1章

あああ… …あ。○

…あああ。○ 』 …段落構造10

と

「第2章

いいい… …い。○

…いいい。○ 』 …段落構造20

というように2つの段落から構成される構造化文書が得られ、これらの段落構造は、構造化文書の論理構造において、兄弟関係となる。

【0152】ここで、変換後の構造化文書を文書部品として出力するように指示されていた場合は以下のように

\*【0144】ステップ1705において、文書の生成の場合は、文書構成要素として、文書構造のルートを示す文書構成要素をファイルに書き出し(ステップ1706)、その後、生成データ一時記憶領域内の情報を全てファイルに書き出す(ステップ1707)。

【0145】一方、文書の生成でない場合すなわち文書部品の場合は、ステップ1707に進む。

【0146】ところで上記ステップ1702においてファイル指定でない場合は、出力先は標準出力であると認識し(ステップ1708)、その後、ステップ1705に進む。

【0147】以上の説明から明らかなように、1つのテキストファイルの内容情報を構造化文書に変換したときは、例えば上述した内容情報1は例えば上述した段落4～6のように分割され構造化される。

【0148】一方、複数のテキストファイルの内容情報を構造化文書に変換したときは、以下ようになる。

【0149】例えば、以下の2つのテキストファイルが指定されたとする。

「第1章

あああ… …あ。○

…あああ。○ 』 …文書部品10

と

「第2章

いいい… …い。○

…いいい。○ 』 …文書部品20

の2つの文書部品が出力される。

【0153】一方、1つの構造化文書として出力するように指示されていた場合は、以下のように出力される。すなわち

「第1章

あああ… …あ。○

…あああ。○

第2章

いいい… …い。○

…いいい。○ 』 …構造化文書10

というように出力される。

【0154】また上述した設定内容において、テキストファイル20の内容情報に対するレベル情報として、レベル「1」「第2章」に代えて、レベル「2」「第1章第1節」が設定されている場合は、上述したテキストファイル10、20の内容情報は以下のように変換される。すなわち

25

「第1章

あああ… …あ。○

…あああ。○」

と

「第1章第1節

いいい… …い。○

…いいい。○」

というように2つの段落から構成される構造化文書が得られ、これらの段落構造は、構造化文書の論理構造において、親子関係となる。

【0155】なお上述した例では、「第1章」や「第1節」としたが、これをタイトル名の指定のときに、第1章を意味する「1」、第1章第1節を意味する「1.1」と設定することもできる。勿論変換後においては「1」、「1.1」となる。

【0156】以上説明した様に第1の実施例によれば、1つの内容情報（テキストデータ）を複数の部分構造に分割することができ、しかも分割した複数の部分構造に構造情報を付加して1つの構造化文書を又は複数の文書部品を得ることができる。文書部品として生成することができるということは、文書部品を再利用することができるということを意味している。

【0157】また処理対象とする内容情報に、特定の構造生成指示記号（例えば段落生成のための記号）を付加する必要がないので、ユーザが、特定の言語や構造化文書の詳細な知識を必要とすること無く、通常のテキストファイルを構造化文書として変換することが可能である。

【0158】更に複数の入力データ（内容情報＝ファイル内の情報又は入力される文字列）から1つの文書又は1まとまりの文書部品として生成することができる。

【0159】本発明の第2の実施例を図18乃至図22を参照して説明する。

【0160】図18は、第2の実施例で使用するGUIシートの構成を示したものであり、このGUIシートは、図4に示したGUIシートの構成に、「段落属性」欄1800を追加した構成になっている。この「段落属性」欄1800においては、スタイル名を入力するためのテキスト入力パーツ、明朝かゴシックかの書体を指定するためのチョイス、文字サイズを入力するためのテキスト入力パーツが設けられている。なお「段落属性」欄1800は論理構造定義手段30の機能を有している。

【0161】図19は、図18のGUIシートにおいてサブパラメータを変えた様子を示したものであり、このGUIシートは、図5に示したGUIシートの構成に、上記同様の「段落属性」欄1800を追加した構成になっている。

【0162】図20は、第2の実施例において、内容情報入力手段10、論理構造定義手段30、出力先定義手段50によって指定された入力情報を記憶するメモリを

26

…段落構造30

…段落構造40

示したものであり、このメモリは、図6に示した第1の実施例のメモリ600の構成に、段落属性指定領域2000を追加した構成になっている。この段落属性指定領域2000には、論理構造定義解釈手段40によって、図18に示したGUIシート400の「段落属性」欄1800に設定された内容が格納されるようになっている。

【0163】第2の実施例は基本的には第1の実施例と同様であるが、異なっている点は、段落属性の処理が加わったことである。

【0164】すなわち、入力情報の解釈処理においては、図9及び図10に示した処理手順と同様の処理が行われ、そして図10のステップ1008の処理においては、図21に示すように、メモリ600の生成段落指定フラグ領域611fに、GUIシート400の「段落生成」欄423に設定されている段落生成の種類を設定し（ステップ2101）、その後、「段落属性」欄1800に段落属性が設定されている場合は、その値を、段落属性指定領域2000に格納する（ステップ2102）。

【0165】変換手段70による変換処理は、図12及び図13に示した処理手順と同様の処理が行われ、また内容情報の出力処理も、図15及び図16に示した処理手順と同様の処理が行われる。文書構造構成用の出力処理については図22に示す処理手順に従って処理が行われる。

【0166】すなわち図22に示すように、変換手段70は、図14に示したステップ1401～1403と同様の処理を行い（ステップ2201～2203）、その後、段落属性指定領域2000の値に基づいて、段落属性があるか否かを判断し（ステップ2204）、存在する場合は段落属性指定領域2000に設定されている段落属性を生成データ一時記憶領域624に出力し（ステップ2205）、その後、構造単位終了記号を生成データ一時記憶領域624に出力する（ステップ2206）。

【0167】ステップ2204において段落属性がない場合はステップ2206に進む。すなわち、この段落属性がない場合の処理は、図14に示した処理手順と同様である。

【0168】以上説明した様に第2の実施例によれば、第1の実施例の作用効果に加えて、元の内容情報の段落属性ではなく、所望の段落属性を指定して、構造化文書

あるいは文書部品を生成することができる。

【0169】本発明の第3の実施例を図23乃至図27を参照して説明する。

【0170】図23は、第3の実施例のGUIシートの構成を示したものであり、このGUIシートは、図18に示した第2の実施例のGUIシートの構成に、「ひな型」欄2300を追加した構成になっている。この「ひな型」欄2300においては、ひな型を使用するか使用しないかを指定するチョイスが設けられている。

【0171】図24は、図23のGUIシートにおいてサブパラメータを変えた様子を示したものであり、このGUIシートは、図19に示した第2の実施例のGUIシートの構成に、上記同様の「ひな型」欄2300が設けられた構成になっている。この「ひな型」欄2300においては、「使用」項目が指定され、この指定により表示されたテキスト入力パーツにおけるファイル名に、「template1.doc」が指定されている。なお「ひな型」欄2300は論理構造定義手段30の機能を有している。

【0172】図25は、第3の実施例において、内容情報入力手段10、論理構造定義手段30、出力先定義手段50によって指定された入力情報を記憶するメモリを示したものであり、このメモリは、図20に示した第2の実施例のメモリ600の構成に、ひな型指定領域2510、ひな型抽出データ領域2520を追加した構成になっている。

【0173】第3の実施例は基本的には第2の実施例と同様であるが、異なっている点は、ひな型の処理が加わったことである。

【0174】すなわち、入力情報の解釈処理においては、図9及び図10に示した処理手順と同様の処理が行われ、そして図10のステップ1008の処理においては、図26に示すように、図21に示したステップ2101、2102と同様の処理を行い（ステップ2601、2602）、その後、GUIシート400（図24参照）の「ひな型」欄2300の値に基づいて、ひな型指定はあるか否かを判断する（ステップ2603）。

【0175】ひな型指定がある場合は、メモリ600（図25参照）のひな型指定領域2510に、「ひな型」欄2300のテキスト入力パーツにおけるファイル名に設定されているファイル名を設定する（ステップ2604）。

【0176】そして、そのファイル名に対応するファイルからひな型となる文書データ（ひな型文書）を読み込み、該文書データ内にスタイル定義情報があれば、その定義情報をひな型抽出データ領域2520に格納する。また「生成構造」欄422に設定されている値（つまり「段落」か「見出し」）に基づいて、「段落」か「見出し」かのフラグを構造情報フラグ領域611cに格納する。更に「段落属性」欄1800に段落属性が設定され

ている場合は、その段落属性を段落属性指定領域2000に格納する（ステップ2605）。これで、入力情報の解釈処理が終了したことになる。

【0177】なおこの第3の実施例においてはGUI400からの生成構造情報及び段落属性指定情報の指示あった場合であっても、ひな型文書の情報を優先させるようにしているが、これに限定されることなく、これらの情報の優先順位は任意に決定することができる。

【0178】このようにひな型文書を参照することにより、このひな型の持つ段落情報を生成する文書又は文書部品に適用することができる。ひな型文書の参照とは、ひな型文書の構造の参照（見出し付きの構造か、単なる段落のみの構造）や、型文書の構造に含まれる段落に設定されている段落属性の参照を意味している。

【0179】変換手段70による変換処理、内容情報の出力処理、文書構造構成用の出力処理は、第2の実施例の処理と同様である。

【0180】最後に、文書化手段90による出力処理について、図27に示すフローチャートを参照して説明する。この出力処理は、基本的には第1の実施例の場合と同様であり、異なる点は、スタイル定義情報の処理が追加されたことである。

【0181】すなわち、図27に示すように、図17に示したステップ1701～1705と同様の処理が行われる。ステップ1705において、文書を生成する場合は、メモリ600（図25参照）のひな型抽出データ領域2520にスタイル定義情報が格納されている場合は、そのスタイル定義情報を出力する（ステップ2700）。このステップ2700を終了した後は、図17に示したステップ1701～1705と同様の処理が行われる。

【0182】なおスタイル定義とは、文書のある部分に定義されている文字段落を修飾する属性（例えば書体、文字サイズなどの属性）を、複数まとめて一つの名前をつけて定義したものである。そして実際に使用するとき、適用したい段落の属性指定位置にスタイル名称を指定することにより、その名称に対応して記憶されている複数の属性が、まとめて、該当する段落に適用される。これをスタイル参照という。

【0183】以上説明した様に第3の実施例によれば、第1及び第2の実施例の作用効果に加えて、型文書を参照可能とし、しかもその型文書の持つ属性情報を、生成する構造化文書又は文書部品に反映させることができる。

【0184】なお、上述した第1～第3の実施例で説明した処理によって得られる複数の文書部品（すなわち複数の段落構造）は、後段で実行する処理に用いることができる。例えば、後段で実行する処理として、本願出願人が先に出願した特願平4-206722号の発明のものによって利用される。



【0185】

【発明の効果】以上説明したように本発明によれば、文書内容が構造化されていない1つの文書から、論理構造を有する1つの構造化文書あるいは1つの構造化文書を構成する複数の文書部品を得ることができる。

【0186】また他の発明によれば、文書内容が構造化されていない複数の文書から、論理構造を有する1つの構造化文書あるいは1つの構造化文書を構成する複数の文書部品を得ることができる。

【図面の簡単な説明】

【図1】本発明に係る文書生成装置の第1の実施例を示す機能ブロック図。

【図2】図1に示した装置を実現するためのハードウェア構成を示す構成図。

【図3】構造化文書を説明するための図。

【図4】第1の実施例のGUI（グラフィカル・ユーザ・インタフェース）シートの一例を示す図。

【図5】第1の実施例のGUIシートの一例を示す図。

【図6】第1の実施例における入力情報が記憶されるメモリの構成を示す図。

【図7】図6に示したメモリ中のワーク領域に設定されるデータの動的な格納処理を説明するための図。

【図8】第1の実施例の概要処理動作を示すフローチャート。

【図9】第1の実施例における入力情報解釈処理動作を示すフローチャート。

【図10】第1の実施例における入力情報解釈処理動作（1番目の領域の入力情報についての処理）を示すフローチャート。

【図11】図10に示した処理手順における「段落に関する処理」のサブルーチン。

【図12】第1の実施例における変換手段による変換処理の動作を示すフローチャート。

【図13】図12に示した処理手順の続きの変換処理の動作を示すフローチャート。

【図14】第1の実施例における変換手段による文書構

造構成要素の出力処理の動作を示すフローチャート。

【図15】第1の実施例における変換手段による内容情報の出力処理の動作を示すフローチャート。

【図16】図15に示した処理手順の続きの出力処理の動作を示すフローチャート。

【図17】第1の実施例における文書化手段による出力処理の動作を示すフローチャート。

【図18】第2の実施例のGUIシートの一例を示す図。

【図19】第2の実施例のGUIシートの一例を示す図。

【図20】第2の実施例での入力情報が記憶されるメモリの構成を示す図。

【図21】第2の実施例における入力情報解釈処理（1番目の領域の入力情報についての処理）における「段落に関する処理」のサブルーチン。

【図22】第2の実施例における変換手段による文書構造構成要素の出力処理の動作を示すフローチャート。

【図23】第3の実施例のGUIシートの一例を示す図。

【図24】第3の実施例のGUIシートの一例を示す図。

【図25】第3の実施例での入力情報が記憶されるメモリの構成を示す図。

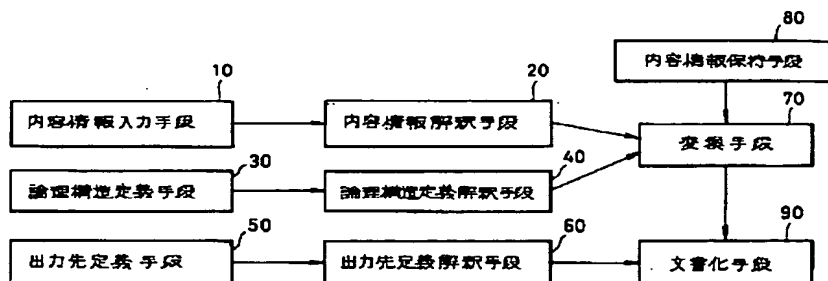
【図26】第3の実施例における入力情報解釈処理（1番目の領域の入力情報についての処理）における「段落に関する処理」のサブルーチン。

【図27】第3の実施例における文書化手段による出力処理の動作を示すフローチャート。

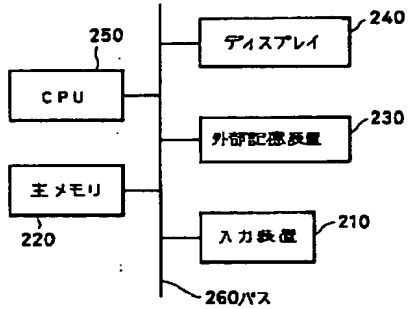
【符号の説明】

10…内容情報入力手段、20…内容情報解釈手段、30…論理構造定義手段、40…論理構造定義解釈手段、50…出力先定義手段、60…出力先定義解釈手段、70…変換手段、80…内容情報保持手段、90…文書化手段。

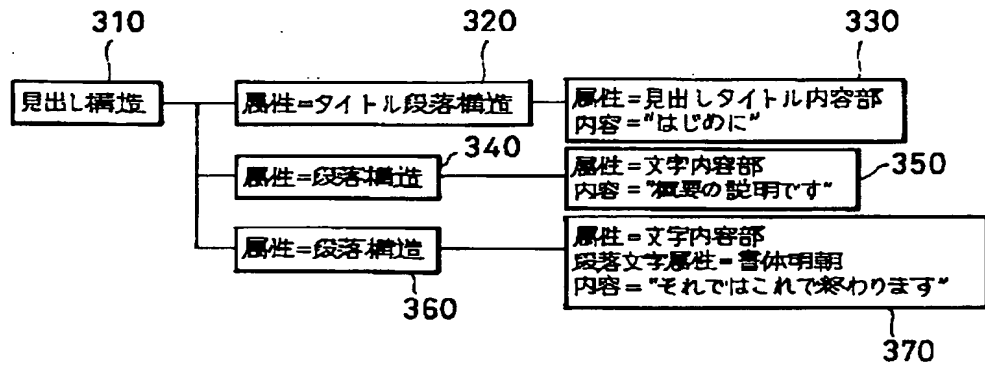
【図1】



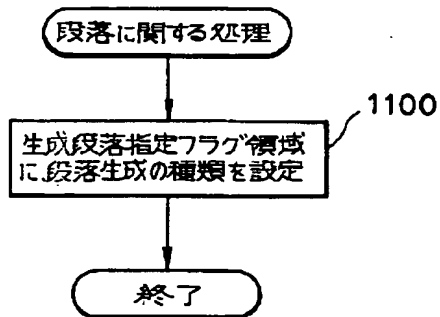
【図2】



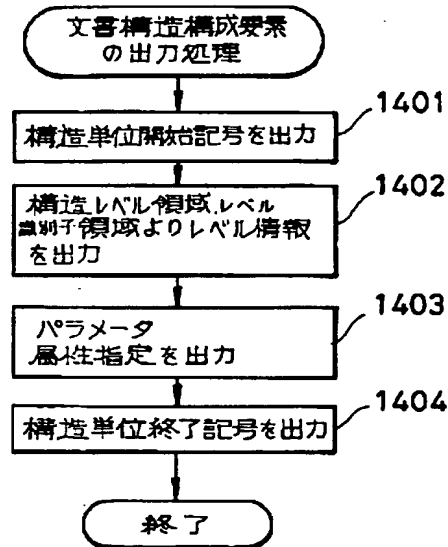
【図3】



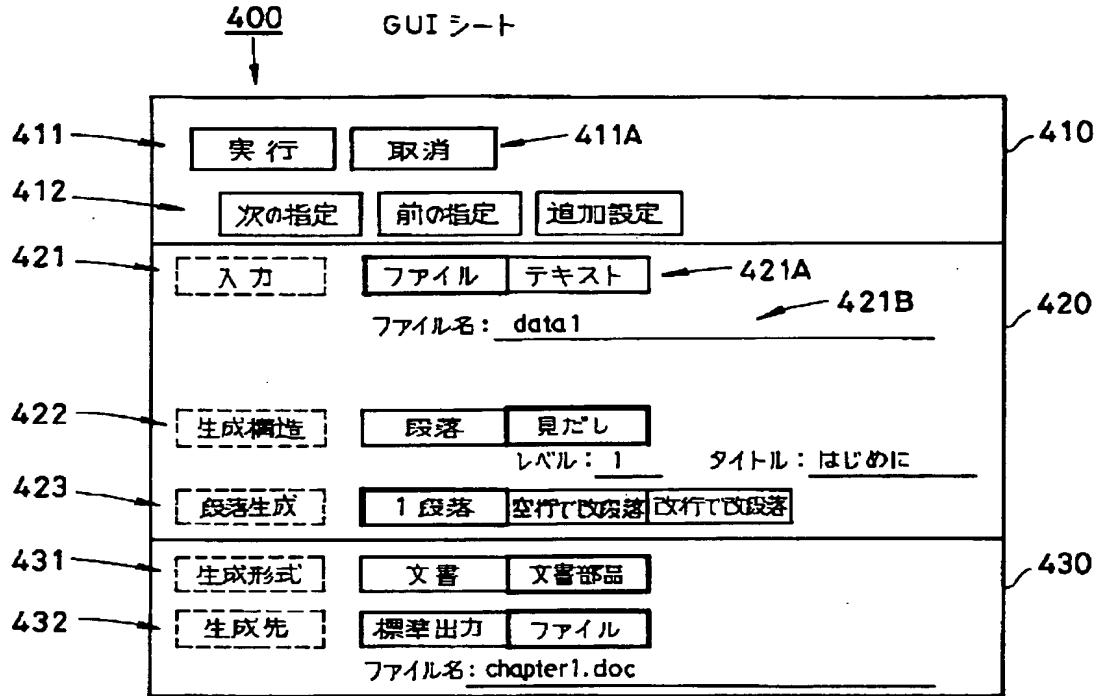
【図11】



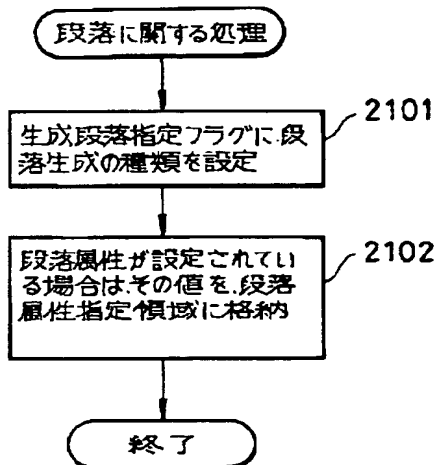
【図14】



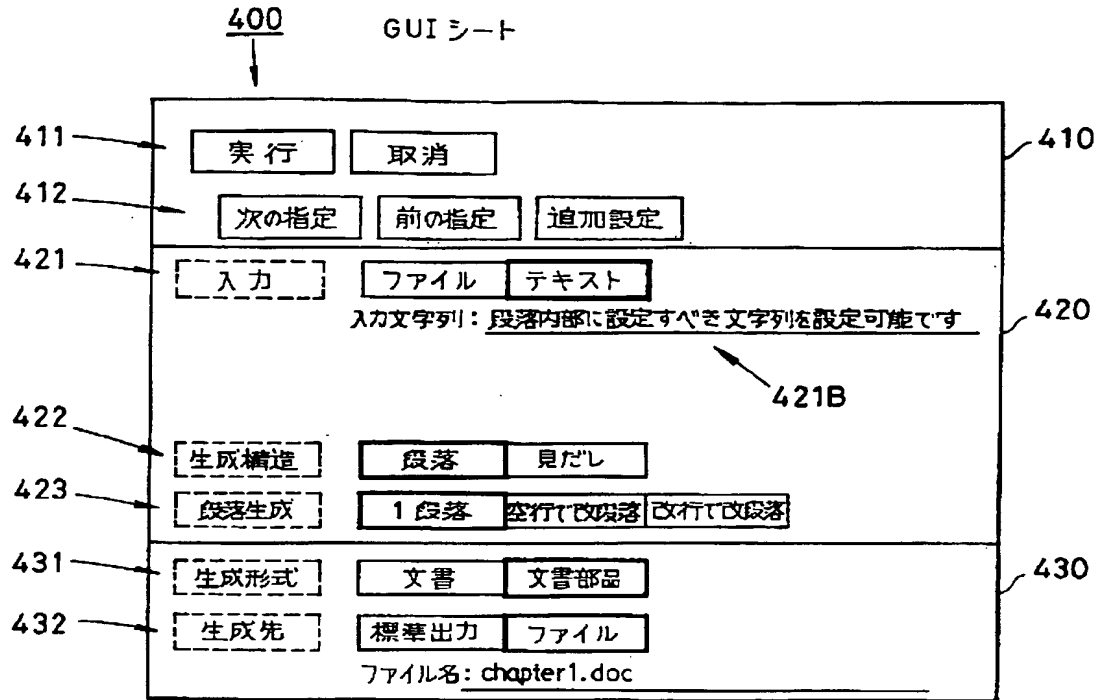
【図4】



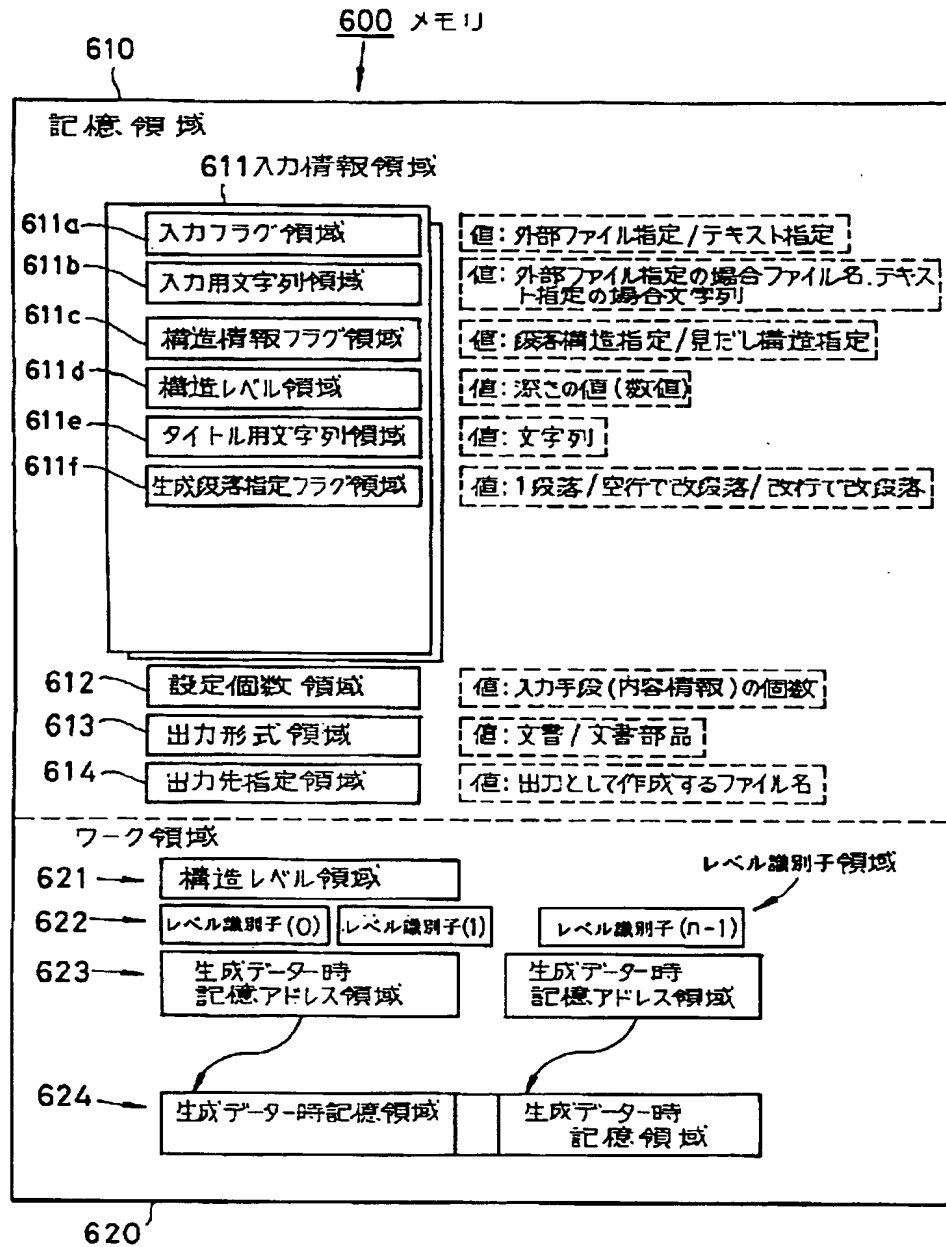
【図21】



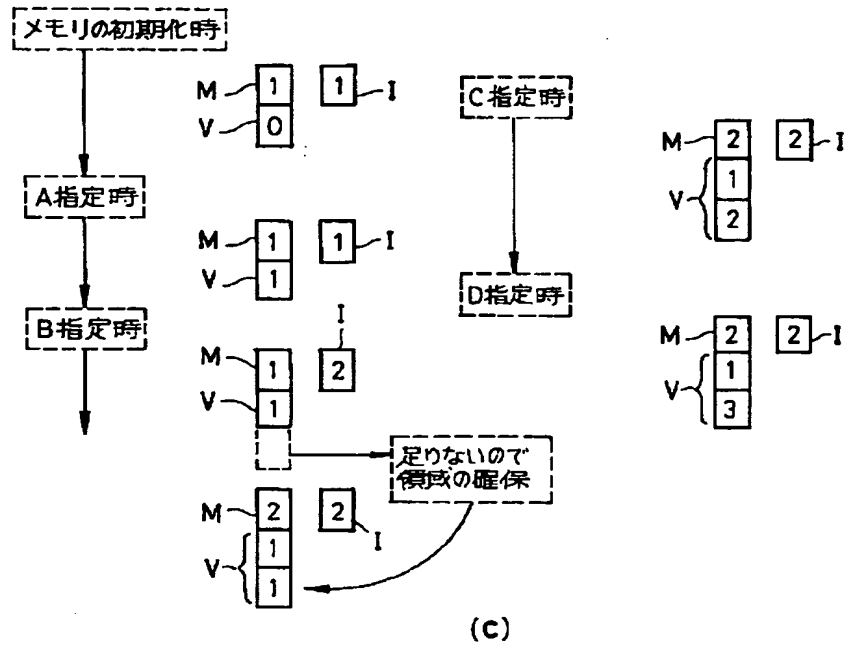
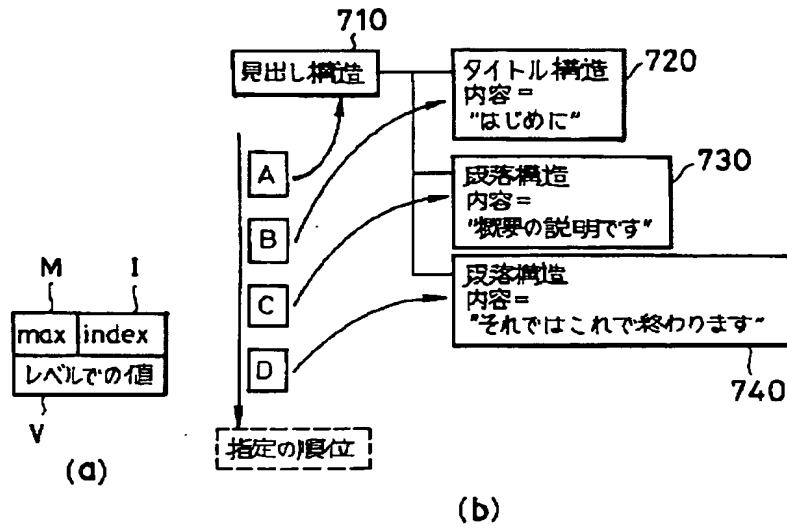
【図5】



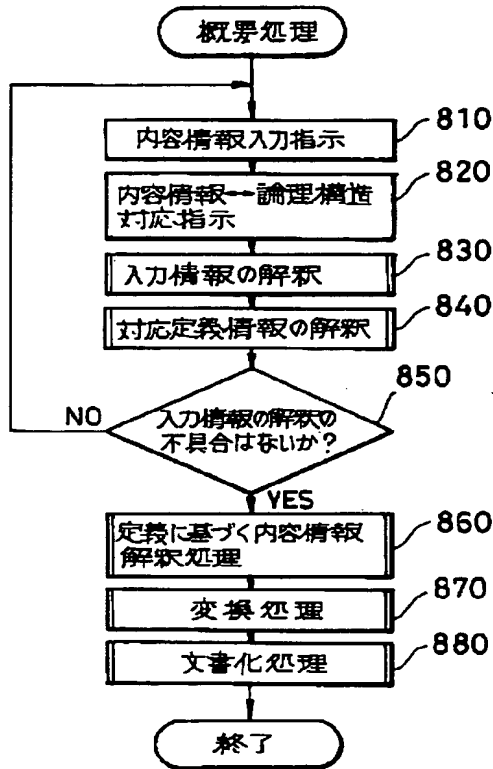
【図6】



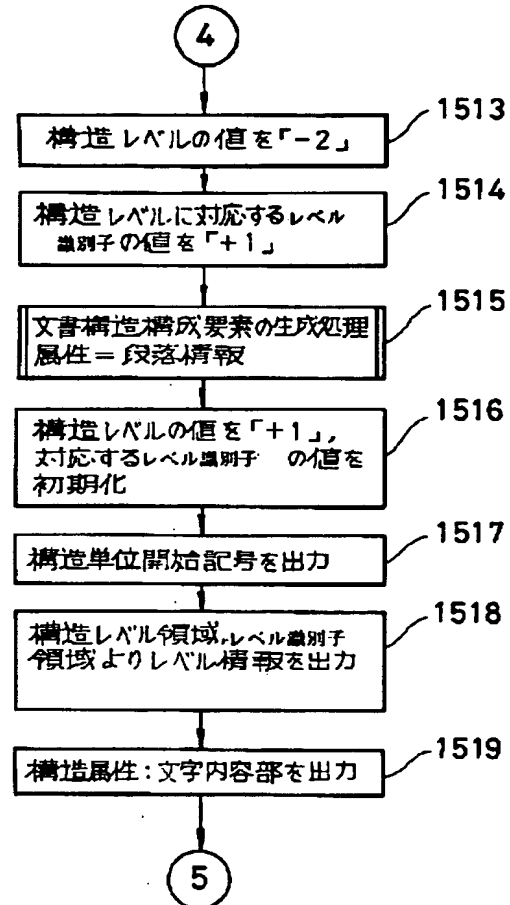
【図7】



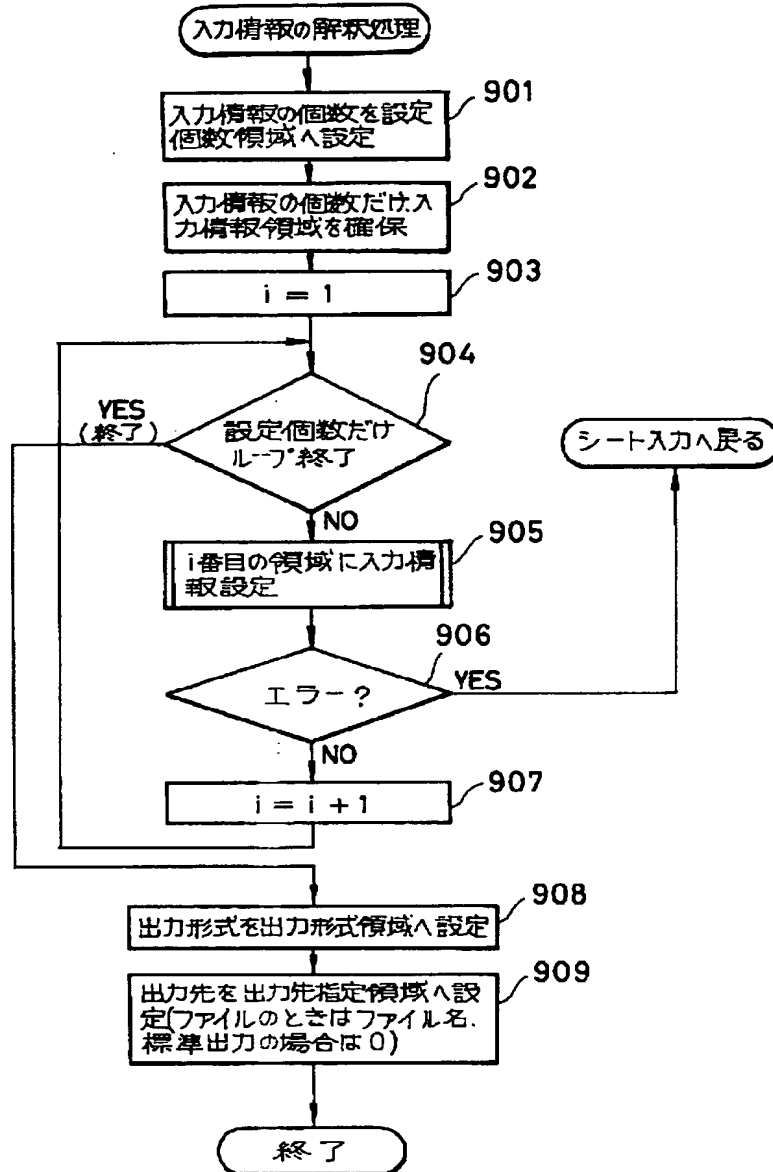
【図8】



【図16】

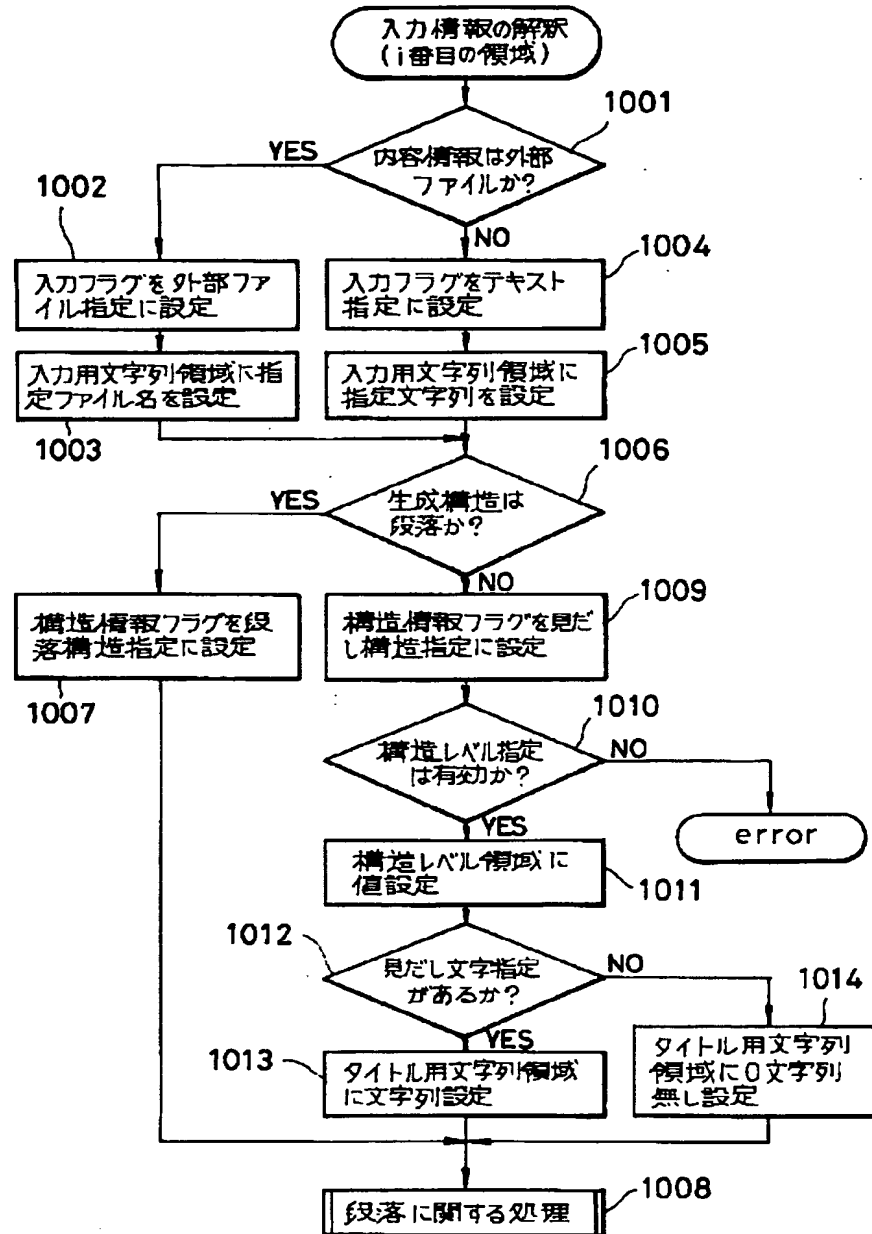


【図9】

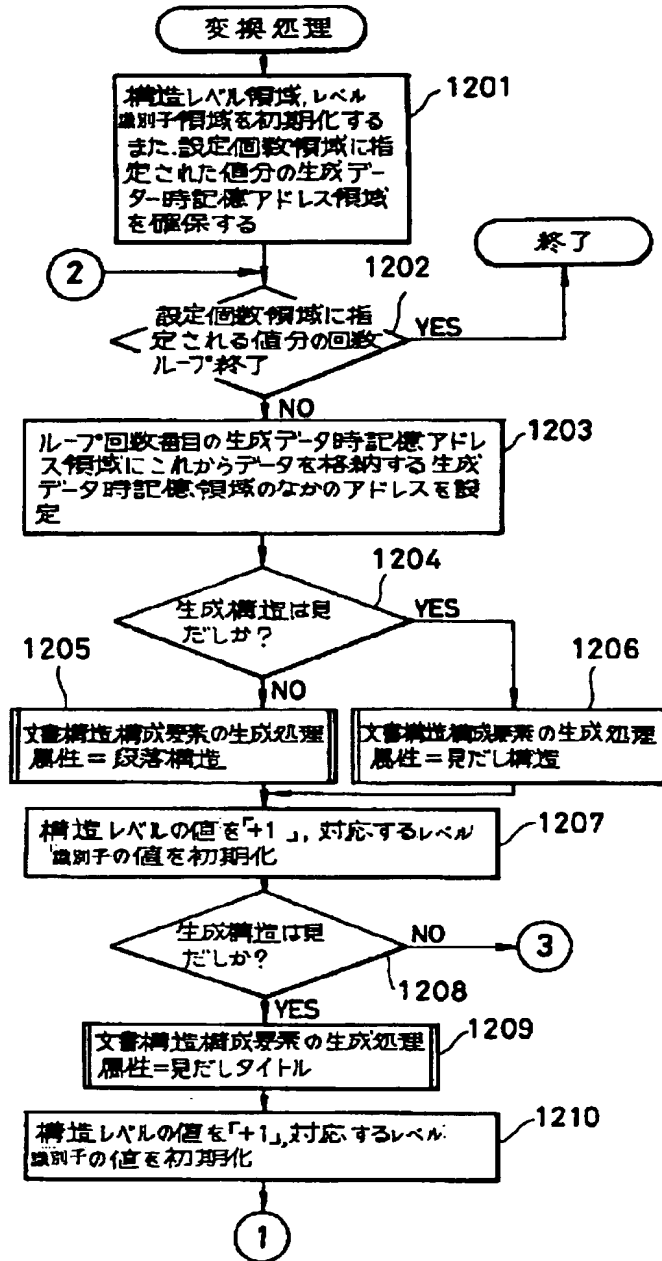




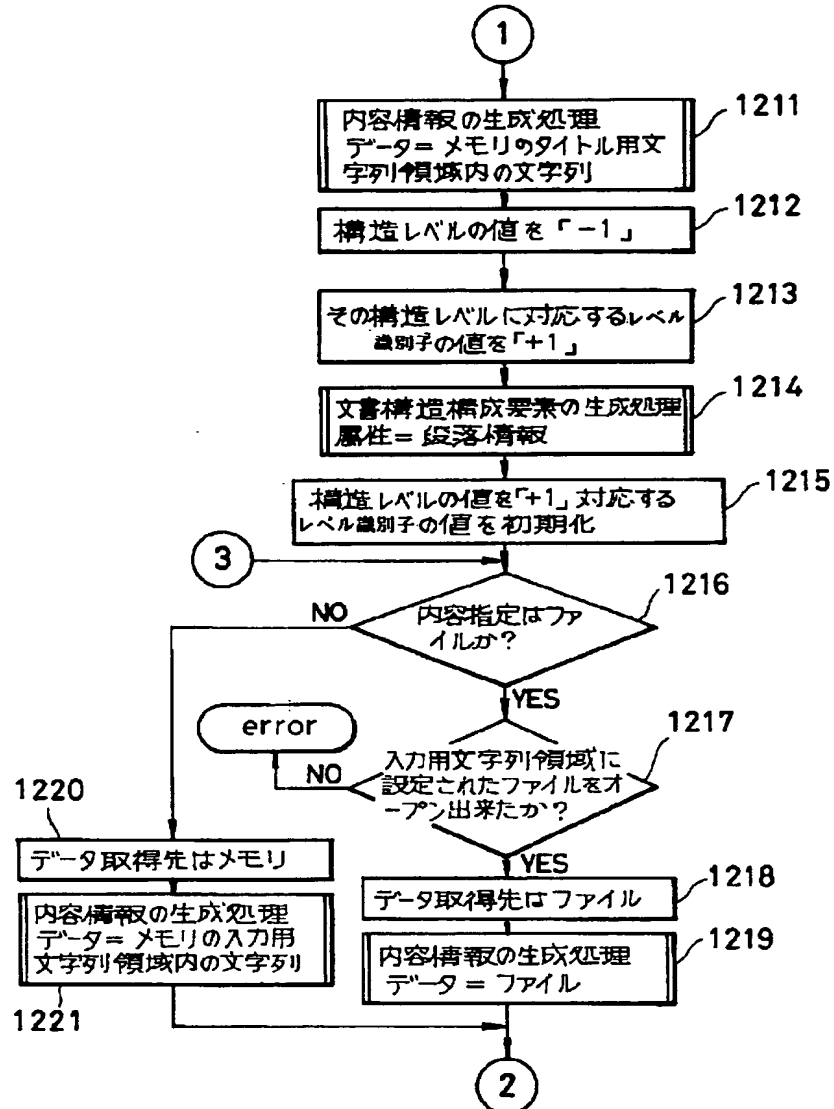
【図10】



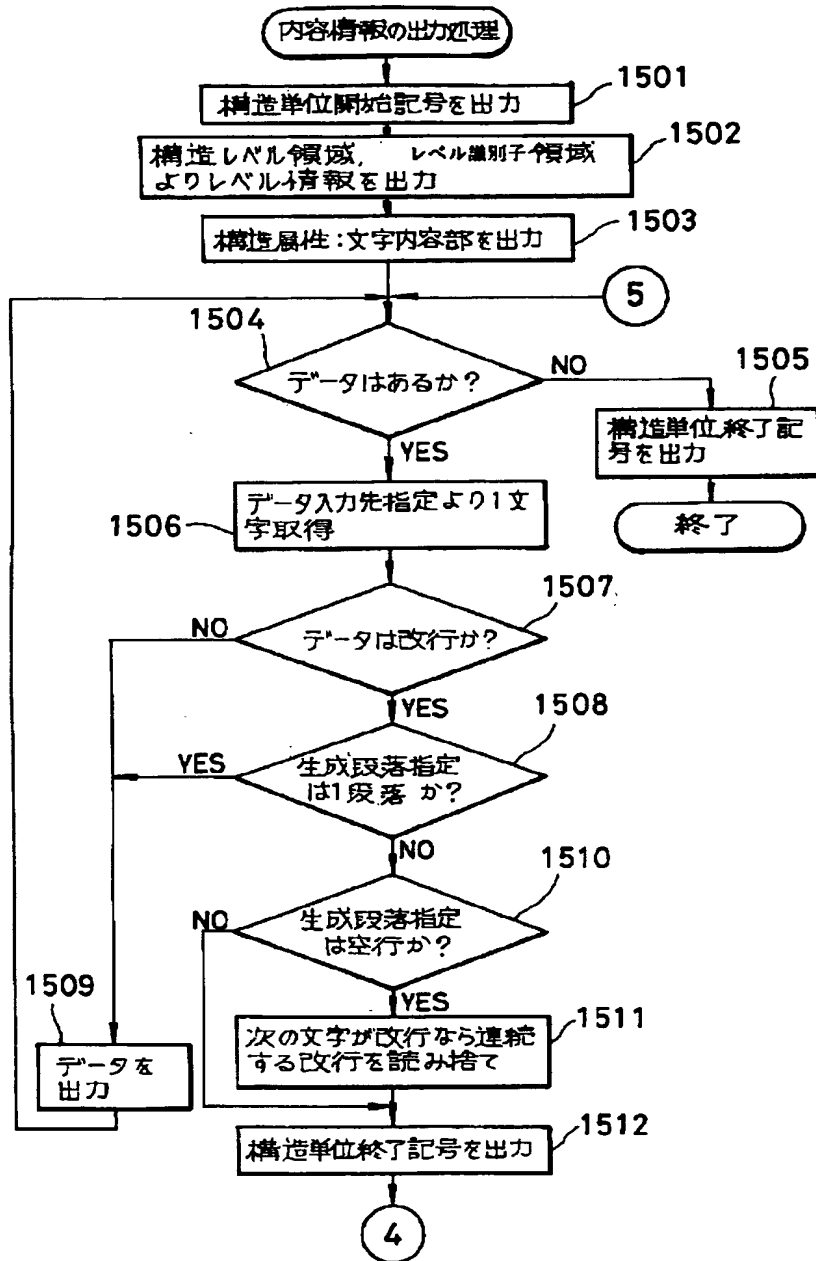
【図12】



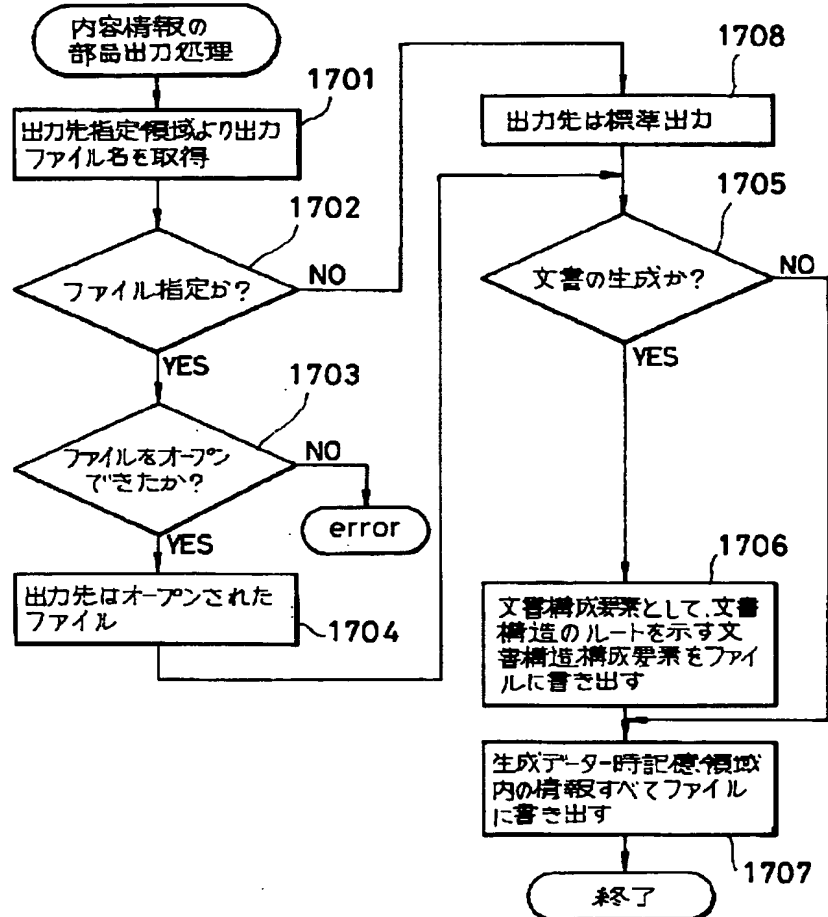
【図13】



【図15】



【図17】



【図18】

400 GUIシート

411 実行 取消

412 次の指定 前の指定 追加設定

421 入力 ファイル テキスト

1800 段落属性

ファイル名: data1

スタイル名: style1

書体 明朝 ゴシック

文字サイズ 12 ポイント

422 生成構造 段落 見出し

レベル: 1 タイトル: はじめに

423 段落生成 1段落 空行で改段落 改行で改段落

431 生成形式 文書 文書部品

432 生成先 標準出力 ファイル

ファイル名: chapter1.doc

【図19】

400 GUIシート

411 実行 取消

412 次の指定 前の指定 追加設定

421 入力 ファイル テキスト

1800 段落属性

入力文字列: 段落内部に設定すべき文字列を設定可能です

スタイル名: style1

書体 明朝 ゴシック

文字サイズ 12 ポイント

422 生成構造

423 段落生成

1段落 空行で改段落 改行で改段落

431 生成形式

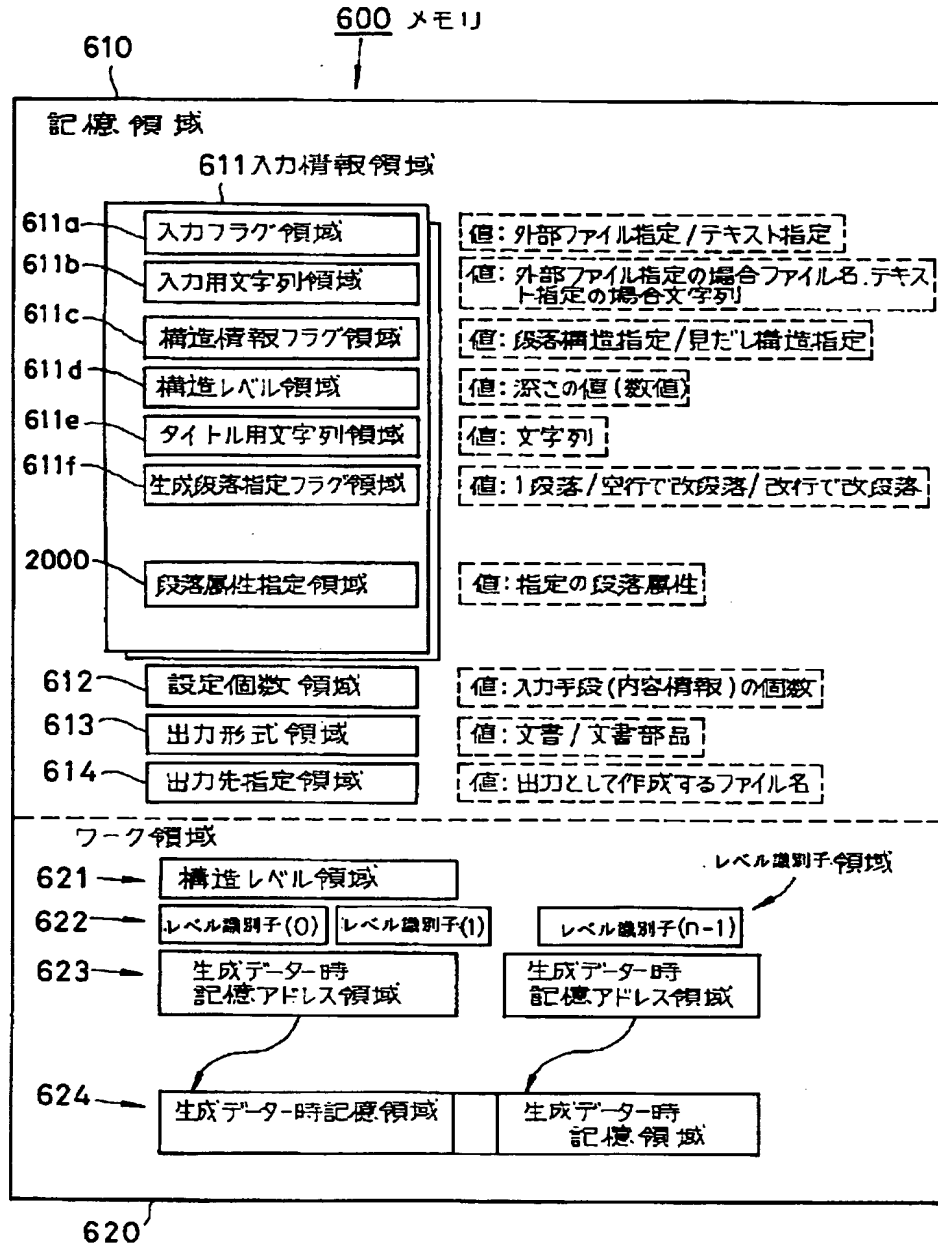
432 生成先

文書 文書部品

標準出力 ファイル

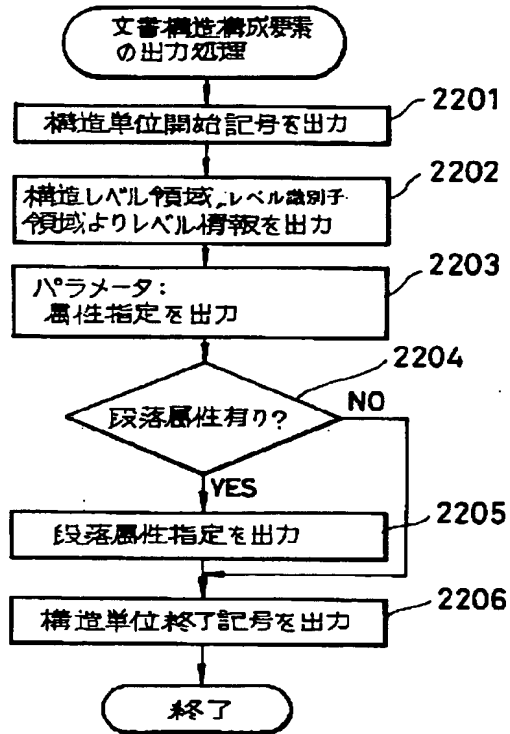
ファイル名: chapter1.doc

【図20】

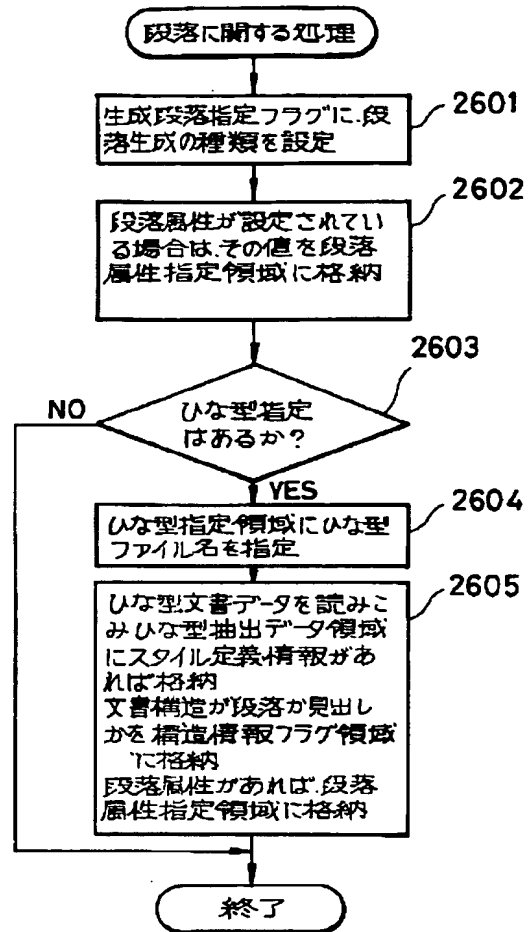




【図22】



【図26】



【図23】

400 GUI シート

411 実行 取消

412 次の指定 前の指定 追加設定

421 入力 ファイル テキスト  
ファイル名: data1

1800 段落属性 スタイル名: style1  
書体 明朝 ゴシック  
文字サイズ 12 ポイント

2300 ひな型 使用する 使用しない

422 生成構造 段落 見だし  
レベル: 1 タイトル: はじめに

423 段落生成 1 段落 空行で改段落 改行で改段落

431 生成形式 文書 文書部品

432 生成先 標準出力 ファイル  
ファイル名: chapter1.doc

【図24】

400 GUI シート

411 実行 取消

412 次の指定 前の指定 追加設定

421 入力 ファイル テキスト

1800 段落属性

2300 ひな型

422 生成構成

423 段落生成

431 生成形式

432 生成先

入力文字列: 段落内部に設定すべき文字列を設定可能です

スタイル名: style1

書体 明朝 ゴシック

文字サイズ 12 ポイント

使用する 使用しない

ファイル名: template1.doc

段落 見だし

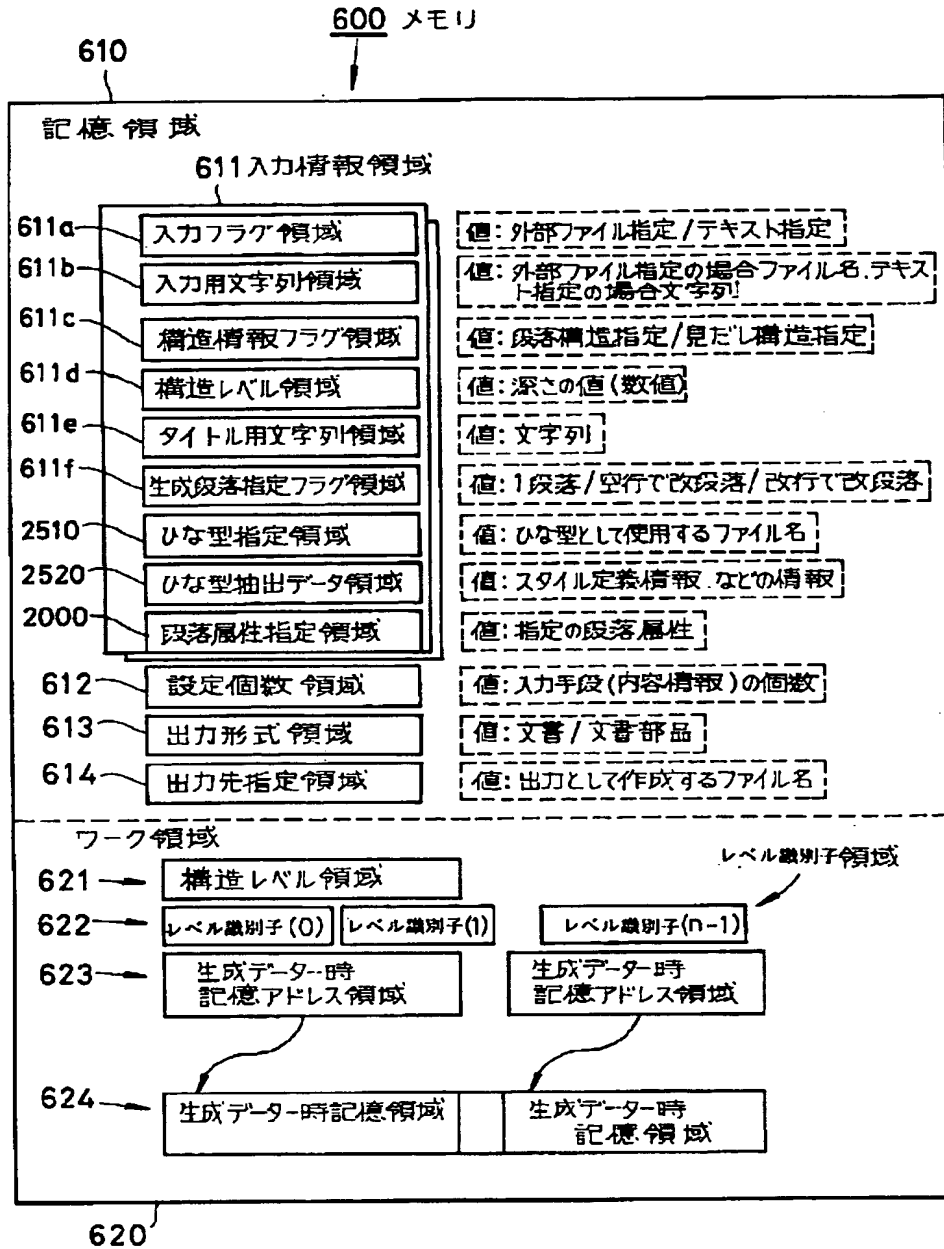
1 段落 空行で改段落 改行で改段落

文書 文書部品

標準出力 ファイル

ファイル名: chapter1.doc

【図25】



【図27】

